DIFFUSION, CO-EVOLUTION AND STRATEGIC INTERDEPENDENCE IN COMPARATIVE AND INTERNATIONAL POLITICS: NEW SPATIAL ECONOMETRIC AND EVENT HISTORY APPROACHES

BY

AYA KACHI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Political Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Associate Professor Milan W. Svolik, Chair
Professor Janet M. Box-Steffensmeier, Ohio State University
Professor Brian J. Gaines
Professor James H. Kuklinski
Associate Professor Tracy E. Sulkin

Abstract

Interdependence is ubiquitous across theories of democratization. For example, the level of democracy in one country might be dependent on its level in other countries; the timing of democratization might be related to the survival of existing democracies. In contrast, much of the empirical literature on democratization has modeled the level of democracy and the timing of regime transitions as if all the observational units and events were independent.

The essays in this thesis explore three sources of interdependence in the study of democratization: the first concerns the causal connection between the emergence and collapse of democracy; the second is the diffusion of political regimes across countries; The third the reinforcement and local convergence of political regimes across countries and over time. Although each type of interdependence raises a unique set of methodological challenges, the emergence of "feedback loops" defines the common mathematical characteristics of these difficulties. A feedback loop is formed when a change in an outcome (e.g., the level of democracy) influences the outcomes of other units, which in turn comes back to affect the outcome that experienced the original shock. Both the inter-event (e.g., the emergence and breakdown of democracies) and inter-unit (e.g., diffusion and reinforcement) dependencies generate these recursive flows of effects across observational units.

In this thesis, I develop two systems of equations (SEQ) models to account for the three sources of interdependence. In all the models presented here, I take a so-called "substantive" approach, rather than a "nuisance" approach, in order to *model* the theoretically-informed structure of dependence. In Essay 1, I develop a multivariate event history model in order

to incorporate the two-way causal relationships between the emergence and breakdown of democracies. In Essay 2, I develop a multivariate event history model for data with right-censored observations, building on the model developed in Essay 1. Essay 3 introduces a new spatial econometric model, which estimates the existence and the strength of both regime diffusion and regime reinforcement, using time-series cross-sectional data of democracy levels.

To my parents, Chiyoko and Satoshi Kachi

Acknowledgements

I am grateful to the broad scholarly community around me, in- and outside of the University of Illinois, for such a positive research experience in graduate school. I would like to extend very special thanks to my advisor, Milan W. Svolik, for his insightful and timely advice both on my research and career, not to mention his patience and encouragement. Throughout the course of discussions on my projects and career, there were a couple of moments when my advisor's comments were, quite frankly, eye-opening and changed my mindset in a useful way. I wish to remember these as I further advance my career in academia. I would also like to thank the rest of my committee members, Janet M. Box-Steffensmeier, James H. Kuklinski, Brian J. Gaines and Tracy E. Sulkin, for their consistent support in pursuing my research. I would also like to thank José Antonio Cheibub for his feedback on earlier iterations of my thesis chapters. Two other scholars who deserve special acknowledgment are Robert J. Franzese Jr. and Jude C. Hays, my coauthors in two articles. The experience of working through the two projects and manuscripts with them helped to shape me as a researcher.

I must also acknowledge William T. Bernhard, who was the Director of Graduate Studies (DGS) when I joined the program, and Tracy E. Sulkin, who is currently the DGS as I graduate from the program, for their timely academic advising. I would also like to thank Brenda Stamm, the Graduate Studies Secretary, who facilitated my life as a student significantly by giving me guidance on day-to-day bureaucracy.

My research also benefited from friendships with many of my fellow graduate students. I owe greatly to Shweta Moorthy and Jason Coronel, in particular. After sharing a lot of time and office space for years, Shweta, by now, knows me in and out. Without her professional and moral support, my graduate school life and research experience would not have been as enjoyable as it turned out to be. The frequent and fun conversations about "scholarly productivity" with Jason Coronel is one of the best memories in graduate school. I am sure we will keep an eye on each other in our future career as well.

I am very much grateful to the Graduate College, the Department of Political Science, the Cline Center for Democracy and the European Union Center at the University of Illinois for providing me with funding opportunities throughout my graduate school life. And lastly, of course, my research very much benefited from the National Science Foundation, who funded me to attend five of the Society for Political Methodology Summer Meetings (although I had to declined once for a health reason), two of the Political Networks Conferences and the Empirical Implication of Theoretical Models (EITM) Summer Institute at UCLA in the summer of 2007. I received invaluable comments from the participants of all of these meetings to improve substantive and methodological ideas in my thesis.

Table of Contents

Chapter 1

Introduction

> Let's face it, the universe is messy. It is nonlinear, turbulent, and dynamic. It spends its time in transient behavior on its way to somewhere else, not in mathematically neat equilibria. It self-organizes and evolves. It creates diversity and uniformity. That's what makes the world interesting, that's what makes it beautiful, and that's what makes it work.
>
> *Thinking in Systems*
> Donella H. Meadows

1.1 Interdependence in Regime Transitions

What determines countries' democracy or autocracy levels? For a long time, comparative scholars have attempted to build empirical models that explain political regime changes. In the literature, scholars have conceptualized and operationalized regime transitions in a number of different ways. Some studies focus on explaining the level of democracy (e.g. measured by the Freedom House and Polity IV democracy scores) given a country and a year. In these studies, each country-year observation is the unit of analysis. Another group of democratization studies explain the timing and occurrence of democratization, or the timing and occurrence of democratic breakdown. These studies operationalize the idea of timing and occurrence in two different ways. The most conventional approach is to explain the latent probability of a country's regime change from one type to another, given a country and a year (e.g., Przeworski et al. 2000). A newer approach is to model directly the spell of time till a transition occurs, either to or from democracy (e.g., Alemán and Yang 2011; Svolik 2008).[1] In these duration analyses, the timing of *democratization* is measured as a spell of time between the emergence of autocracy and the transition, or between the independence of a country and transition. Likewise the timing of the breakdown of democracies is measured as a spell of time between the emergence and collapse of a democracy.

In this thesis, I highlight the inadequacy of existing empirical studies, by revisiting and restructuring three important theoretical arguments ignored in existing research. In a nutshell, the common problem lying beneath these empirical approaches is the assumption of *independent observational units*, when the units are highly interdependent in reality. First, I point out that the emergence and breakdown of democracies are interdependent, while existing studies look at each event in isolation. The second type of interdependence is the contagion of political regimes across countries. Diffusion studies exist, but most of their empirical approaches reduce the inter-state relational information down to country-specific attributes. The resulting empirical models can only test, at best, whether the average surrounding democratic/autocratic environment affects a country's democracy level. These models cannot test whether there exist *influence* dynamics among countries and how strong these influences are between each given pair of countries. Finally, I develop a theory of the regime-support dependency. I argue that countries are not necessarily "open to change" their regimes as the conventional diffusion concept assumes. They can also attempt to maintain their current democracy/autocracy levels, for better or for worse. In this case, countries seek peers that have currently similar regimes and form political/diplomatic ties with them. Through this regime-support and regime-approval network, countries can influence each other in a way that they reinforce their regimes over time. Once the regime-support dependency is de-

[1] I should note that Svolik (2008) mainly contributes to distinguish empirically a subtle difference between cases where countries do not seem to democratize further because they are already consolidated democracies, and cases that are still at risk of reverting to authoritarian regimes. Hence the concept behind the dependent variable is not simply about "democratizing or not". However, this still represents a study of democratization, broadly defined, that operationalizes transitions as spells of time from one state to another, or equivalently, failure rates of democracies.

fined, the influence mechanism across country is similar to the conventional diffusion theory in a cross-section. However, I need to separate the two—the conventional contagion and regime-reinforcement mechanisms—because the long-run dynamics generated by them are very different. This is due to the fact that regime-support networks are defined by similarity in the very outcome that the model tries to explain; i.e., political regimes.

Each of these three sources of interdependence creates its own methodological problems for empirical analysts, but there is a common mathematical characteristic lying underneath all the methodological problems; which is the notion of *feedback loops* in systems theory. At the abstract level, feedback loops are the main reasons why one needs more complex tools to make statistical inferences with the presence of interdependence among units. In the following, I elaborate on the (i) theoretical source of interdependence and (ii) the type of feedback loop generated by the theory, for each of the three types of interdependence in democratization. At the end, I lay out the plan of the thesis, focusing on what methodological solutions I suggest to each problem.

1.1.1 Interdependence (1): The Emergence and Breakdown of Democracies

Theoretical works on democratization suggest causal connections between the emergence and breakdown of democracies. Conventional empirical studies explain *either* the timing of transitions to democracy *or* the survival of existing democracies in isolation. However, there are a number of untested theoretical works that suggest (i) the speed of democratic transition depends on how likely the future democracy can be compromised (e.g. by a coup by the former authoritarian elite), and (ii) the timing of breakdown of democracy might depend on how and how quickly the democratic transition occurred.

For example, actor-oriented theories (namely game theoretic models) of democratization (e.g., Acemoglu and Robinson 2006*a*) claim that, in deciding whether to democratize or not, authoritarian leaders might be aware that, later, staging a coup to reverse the newly-established democracy is possible. This anticipation by the elite about the survival of the potential democracy affects their decisions on when to democratize the country. If they expect that it would be costless to reverse the democracy by a coup, the elite might not negotiate hard with the democratic opposition, making the transition process quicker. On the contrary, if they expect that the democracy lasts longer once it emerges, then they might negotiate harder with the opposition, prolonging the transition duration as a consequence. Østerud (2011) also provides a theoretical argument about the elite's anticipation, based on cases of post-communist regimes in east and central Europe. He claims that authoritarian leaders are more willing to negotiate with the democratic opposition, when "they conceive this as a lesser or less risk evil"; in other words, authoritarian leaders are more willing to negotiate with the democratic opposition, when they sense that they might be able to retain some political power and influence after the transition. A specific example would be "the potential cost of coup" that the elite could stage to regain political power. This strategic thinking by the authoritarian elites can induce a positive causal effect from the expected duration (durability) of a future democracy to the length of the current dictatorship. The longer (expected) survival "causes" a prolonged dictatorship period.

It is not only the elite that is concerned about the future regime outcome. The democratic opposition can also forms anticipation about the success and effectiveness of its anti-government movements. For example, citizens living in dictatorship attempt to gauge how likely a collective action against the authoritarian government would succeed. They might simply sense the odds of success, or they might learn from an experience of neighboring countries (Weyland 2009). The movements can also occur within the government. The opposition force within the government might gauge how much of political power (relative to the authoritarian incumbent) they possess to push for a new democratic constitution. For example, Seely (2005), comparing Benin's successful case and Togo's unsuccessful case of democratization, argues that the pro-democracy group's (including both governmental and civic actors) relative political power facilitated Benin's democratization. However, in my view, it is also important to note that the democratic opposition does not usually confront against the incumbent authoritarian regime, just because they know they have greater political power relative to the incumbent, because if the liberalization movement or negotiation does not lead to a lasting democracy in the future, they would most likely be punished by the continuing authoritarian government. This is a vital component of the classical collective action dilemma with the presence of severe punishment. The same applies to the theory of economic crises and democratization. Gisselquist (2008), for instance, partially attributes Benin's successful liberalization to the increasing pro-democracy movements in the late 1980's. She argues that those movements and the National Conference that eventually led to a new democratic constitution were triggered by economic crises. To me this is only half convincing. It is because, again, the opposition would not participate in costly collective actions unless they anticipate a higher probability of success in the future. In fact, Nigeria in the late 1980's was also a dictatorship that was experiencing a major recession, but liberalization in Nigeria did not occur for another decade. In each of these cases, unobservable strategic calculations and assessment by the opposition involved their expectation about how likely their costly political actions would lead to the emergence and ideally the stability of a democracy.

Take another case in the Eastern and Central European countries in the late 80's and early 90's. After five years of Mikhail Gorbachev's leadership in the Soviet Union, the direct rule of the Soviet communist party came to an end and the effect of Soviet power started to decline even in other east European countries. With the stagnating economy and the weakened political authority of the centralized communist regimes, the liberalization demand from these societies heightened. There were increasing number of major and minor uprisings by the opposition. Lewis (2000) points out that gain in the political power by the liberalization forces was the key in the course of this change in the political atmosphere. He claims that the relaxation of dictatorship and repressive practices after Stalin's death in 1953 was not enough of liberalization of the regime, but it was "sufficient to permit freer communication and a degree of contact with the west that only made awareness of relative failure of the communist system that much sharper." This awareness of the economic and institutional failure of the communist regime in general invigorated the opposition force toward the end of the 80's. We, researchers, cannot observe exactly what triggers the opposition's "awareness" for the promising political atmosphere, but the key political dynamic one should note is that, to some extent, the opposition can sense the expected effectiveness of their movements and the potential emergence of a democracy in the future, and it affects the actual occurrence of revolutionary uprisings, shortening the duration of the dictatorial regime. To summarize the essence of these cases in Africa and Europe, the longer (expected) survival of democracies can "cause" a shorter dictatorship period through heightened political activities by the democratic opposition. Note that the causal direction is the same as the first case, but the effect is opposite.

Finally, there is also a dynamic that manifests the opposite direction of causal relationships. Gradualism advocates claim that quick transitions to democracies can lead to unstable (unconsolidated) democracies if transitions occur under premature socio-economic conditions. Mansfield and Snyder (2007, 1995, 2005), for example, argue that introducing multiparty elections under authoritarian regimes would more likely lead states to civil wars rather than stable democracies, due to the premature institutional foundations. This claim suggests a causal effect *from* the timing of the emergence of democratic institutions *to* the timing of their breakdown.

Almond and Verba (1965), Putnam (2002) and Moore (1966) imply a similar causal relationship between the emergence and breakdown of democracy, but their arguments emphasize the positive effects of gradual democratization on the stability of future democracies. Moore (1966) argues that incremental inclusion of different classes in politics, accompanied by the economic development and the development of liberal socio-political culture led Britain to a long-lasting democracy. Almond and Verba (1965) and Putnam (2002) argue, from more behavioral perspectives, that a gradual transition to democracy is favorable for the later stability because it cultivates the civic cultures that potentially prepare the citizens to value democracy.

Once we incorporate the theories of two-way causal relationships between the emergence and collapse of democracies, the "independent-unit" assumption made in traditional empirical analyses becomes questionable. Methodologically, what challenges the independent-unit assumption is a feedback loop—a flow of causal influences that go back and forth among the multiple events. A feedback loop is formed when a change in an outcome influences the outcomes of other units/events, which in turn comes back to affect the outcome that experienced the original shock. Suppose that the duration of democratic transition becomes shorter in a certain country than what it would have been. (Think of this as an external democratic shock, for the purpose of a thought experiment.) This is a change in an outcome event, the emergence of democracy. This change—a shortened democratization duration—might be associated with premature socio-economic conditions, leading the newly established democracy to an unstable regime in the future. This generates a flow of causal effects from the timing of democratization to the prospect of consolidation (i.e. the timing of the breakdown). The other causal relationship emerges in the elite's strategic thinking at the beginning of the democratization process. As Acemoglu and Robinson (2006*a*) suggest, an authoritarian elite could form an anticipation about the durability of future democracies, or similarly the cost of staging a coup in the future to reverse the democracy. Elites' anticipations about the survival of future democracies affect their decisions on when to democratize the country. If they expect that it would be costless to reverse the democracy by a coup, the elite might not negotiate hard with the democratic opposition, making the transition process quicker. On the contrary, if they expect that the democracy lasts longer, then they might negotiate harder with the opposition, prolonging the transition duration as a consequence.

This strategic thinking by elites generates a causal effect from the survival to the emergence of democracies. As such, the two durations of our interest influence each other recursively, generating feedback loops.

If we suspect the two-way causal relationships exist, we should not model the timing of democratization or the odds of democratic consolidation by focusing only on one of the two phenomena, because it would not take into account the feedback loops that feed into the two political outcomes recursively.

1.1.2 Interdependence (2): The Contagion of Democracies

Another important source of interdependence in democratization is the regime influence across countries. Since geopolitical studies led by O'Loughlin et al. (1998) provided the evidence for geographical clusterings in countries' democracy/autocracy levels, scholars have attempted to model, theoretically and empirically, what might cause the seeming clusterings of regimes. This is where the theory of diffusion emerged. They theorized that a country's democracy level can influence others', just as a flu and habits travel across individuals who are connected. If citizens in these countries learn about the odds of successful revolution against the authoritarian government by observing the masses' behavior of other countries, revolutions might "spread" around the world, possibly turning such uprisings into transitions to democracies. Studies show that citizens do learn rationally from successful revolutionary movements in other nations (e.g., Weyland 2009). First, scholars suspected geographical proximity as a main pathway that "connects" countries. (e.g., Gleditsch and Ward 2006): countries that are geographically closer affect (or learn about) each other more than those farther away. Later, Beck, Gleditsch and Beardsley (2006) pointed out that "space is more than geography", suggesting a possible effect of implicit economic ties among countries as pathways of democratic diffusion. Goodliffe and Hawkins (2011) took a step further and considered three non-geographical dependency networks (trade, alliance, and international organization (IO) partners) through which democracy and autocracy might spread across countries.

Various other empirical studies of diffusion have been developed since then, but the important theoretical theme that commonly lies beneath all these theories is the idea that models that explain each country's democracy/autocracy level by its own domestic characteristics is not a sufficient representation of democratization. In other words, the independent-unit assumption is violated. Again, methodologically speaking, the violation of the independent-unit assumption stems from feedback loops of democracy levels generated by the diffusion flow.

To see how a shock given to the outcome of a certain observational unit comes back to its own outcome again, imagine the following flow of regime influence. Suppose country A's democracy level ("outcome") rises after a successful wave of social movements. By observing the successful riots in country A, democratization movements in country B become more organized and effective, leading to an increase in country B's democracy level (the outcome of another observational unit). The flow of diffusion does not have to stop at this point. Democratization in country A can further progress after their citizens' observing country B's success. Alternatively, if the democratization movement in country B fails due to lack of their prerequisite political-economic conditions, for example, then this might now slows down movements in country A. In either case, the original change in the democracy level in A has come back to A's own political outcome, after influencing the outcome of another observational unit. This is a feedback loop.

When we believe that these influence loops occur all simultaneously, we cannot predict a country's level of democracy only by its own political economic attributes for a single unit, because it would not take into account the feedback that involve other countries' democracy levels. This simultaneity occurs more often than one might imagine. Since typical data in comparative and international politics are aggregated at the annual level, these dependencies can easily occur "simultaneously" within a single time point of the given data; i.e., it does not have to be literally "instantaneous". In fact, it is known that when researchers overlook the spatial dependence (when it actually exists in data), the estimated effects of country-specific determinants of democracy levels will be biased in favor of the strength of the country-specific covariates (Franzese and Hays 2007); i.e., the effects will be *overestimated*.

1.1.3 Interdependence (3): Self-Selection and Regime Reinforcement

conventional diffusion studies summarized above offer insightful theories about the democratic (or authoritarian) diffusion; and yet, in my view, even their definitions of space is limited.

The third source of interdependence that I introduce is regime reinforcement across countries and over time. Part of this mechanism is similar to that of diffusion, but the reinforcement theory adds the inter-temporal connection to the conventional concept of contagion, which occurs across counties.

As the traditional diffusion studies theorize, I also recognize that countries affect each other's regimes such that they change their institutions toward the direction of others' regimes. However, I theorize that, at the same time, countries also attempt to maintain actively their current regime types. I introduce the idea of the regime-support or regime-approval dependency among countries; i.e., states *select* themselves into a dependency network, where the connections are defined by countries' regime similarity. Country networks of support and approval is difficult to observe, but, for example, Brinks and Coppedge (2006) argue that political leaders use their neighbors' regimes as "good or bad examples" in favor of their own regimes. Neumayer (2008) shows countries tend to form stronger diplomacy ties when their political ideologies are more similar. For instance, in December 2011, the Brazilian Embassy in Chile disclosed telegrams sent to the Chilean governments during the Pinochet regime (Pedigo 2011). They revealed that the Brazilian military dictator Emílio Médici granted a fifty million US dollar worth of economic aid to the newly-emerged Pinochet government. The telegrams also reveal that, at one point, the Pinochet government explicitly asked the Médici government for its political support in front of the international community. These examples suggest that countries form diplomatic support/approval networks precisely when they have more similar—than dissimilar—regimes, selecting themselves into a regime-support network. Hence the strength of ties (or "edges", in the SNA language) of such regime-support networks should be defined by the degree of similarity in their democracy levels from the previous time period.

The regime influence through the support network generates even more complex feedback loops than the conventional diffusion dynamic. This is due to the fact that regime-similarity networks emerge from *homophily*; i.e., the dependency ties, which provide the platform of diffusion, are defined by the very outcome—democracy levels—of the model. Homophily is a phenomenon in which individual units form connections with similar others. McPherson, Smith-Lovin and Cook (2001) extensively review studies that demonstrate homophilic tie formation among individuals, which include situations where one becomes friends with others due to their similar behavioral habits (e.g., smoking, preferring to work in a certain industry etc.). The formation of regime-reinforcement network is a homophilic dynamic in which states form dependency ties with others when their regimes are similar.[2]

With the presence of regime-support networks, the emerging feedback loops is more complex than what we saw for simple diffusion. We can trace the feedback loops as follows. First, countries with more similar regimes form stronger dependency ties of political (diplomatic)

[2]Networks defined by geographical proximity, on the contrary, are completely fixed regardless of countries' preferences. Likewise, economic networks are defined by numerous factors that are outside state similarity in any aspects; for example, obviously trade networks depend largely on the necessity for certain goods and services, geographical distance, and historical incidents that facilitate interstate transactions such as colonialism. None of these are purely correlated with the similarity in countries' political, economic and social aspects.

and economic support. Second, through this dependency network, the contagion of regimes occurs. In this case, however, contagion or influence implies confirmation of each others' regimes, rather than a change from the current regimes. With these updated democracy levels, an updated support network emerges and another wave of the reinforcing influence occurs in the next time period. In the long run, as the consequence of self-selection (into support networks) and diffusion (through such networks), countries' political regimes can be reinforced. In this thesis, I refer to this dynamic as *network-behavior coevolution*: the latent connectivity across countries ("network") and their democracy level ("state behavior") co-evolve over time. With my theory of regime reinforcement, feedback loops emerge not only across countries in each time period, but those contagion dynamics are tied intertemporally as well. With the presence of both the conventional contagious dynamic and the possibility of reinforcement over time, it should be obvious that a model that explains the level of democracy only for each country-year is not a preferable representation of what we think is democratization.

1.1.4 New Statistical Tools: "Substantive" Approaches to Account for Interdependence

The methodological innovation of this thesis is to provide two sets of new statistical tools that account for the three sources of interdependence in democratization. Essay 1 investigates the dependence between the transition and survival of democracies with a new multivariate duration model. Essay 2 extends the multivariate duration model such that it properly accounts for democracy spells whose end-points are not observed within the data time-frame (*right-censoring*). Essay 3 introduces a new spatial econometric model that not only evaluates the contagion of regimes across countries, but also self-selection and reinforcement of regimes over time.

This is not to claim that there has never been an empirical strategy for inter-event, spatial or spatio-temporal dependencies in data. On the contrary, I share the concern with them that researchers should always address such dependencies when they can. However, much of existing work treats the dependence as a statistical "nuisance" rather than a part of political stories to be explained. For example, Boehmke (2006) models the dependence in the timing of Congresspersons' issue position taking, using a seemingly unrelated (duration) regression (SUR) model. The author substantively theorizes that strategic behavior by Congresspersons induces the interdependence among their position-announcement timings. However, with the SUR model, one can test for the *existence* of the dependence at best, and fails to test for the *structure* of the dependence informed by the theory. Qualitatively, this is equivalent to "controlling for" dependencies statistically, but not interpreting the dependencies substantively.

A similar argument applies to various error correction methods widely used in applied political-science research. These include the feasible generalized least squares (FGLS) estimator by Parks (1967), and the robust panel-corrected standard errors (PCSE's) estimators provided by Beck and Katz (1995) in order to account for temporal and spatial dependence in panel and time-series cross-sectional (TSCS) data. Since Beck and Katz (1995), the vast majority of applied empirical studies with panel and TSCS data have reported the robust PCSE's. Similarly, Beck, Katz and Tucker (1998) suggest splined time in logistic regressions to account for time dependency in binary data.

A common problem with the above approaches is that since researchers treat the dependency as a "nuisance," and are so focused on "correcting for" or "controlling for" the dependence in data, they tend to ignore it when it comes to interpreting the regression results as if the dependence did not exist (Franzese and Hays 2007; Carter and Signorino 2010). Similarly, Beck and Katz (1996) also address this problem and suggest that researchers model temporal dynamics by time-lagged dependent variables instead of by the error structure. They maintain "[t]he lagged dependent variable approach makes it easier for researchers to examine dynamics and allows for natural generalizations in a manner that the serially correlated errors approach does not."

In this thesis, I subscribe a different approach—a "substantive" approach—to account for interdependence. I maintain theoretically-informed structures of dependence in all of my empirical models. For example, in Essay 1, I develop a simultaneous equations model for duration variables, such that parameters associated with durations on the right-hand side of equations evaluate the degree to which durations influence each other. This modeling approach allows me to assess and interpret the relative importance of the duration dependence, *as well as* the effects of other structural determinants of each duration—transition and survival. Similarly in Essay 3, I *model* both the contagion and reinforcement dynamics as informed by substantive theories. This allows me to assess and interpret the relative strength of the dependencies *as well as* the effects of other structural determinants of democracy levels.

In addition, in terms of the inferential accuracy, the presence of system dynamics give rise to the following problems when we employ traditional single-equation approaches as in existing empirical studies. When there is dependency among observational units (e.g. the interdependence between the emergence and breakdown of democracies, or contagion and reinforcement of political regimes across states and over time), the estimated effects of the country-specific variables from traditional regression approaches would give us misleading results. The estimates would suffer from omitted variable biases if the dependency is completely ignored, and they suffer from simultaneity biases if we naively use the occurrence of one event as one of the predictors of the other, in a single-equation framework. For example, if a spatial dependence exists in data but is omitted from the model, it is known that the effects of other structural (or domestic, country-specific) variables will be *over*estimated. In other words the empirical finding will be biased toward finding stronger effects of structural determinants (Franzese and Hays 2007).

In this thesis, I present a solution by developing two kinds of systems of equations models (SEQ). The two models respectively accommodate the particular data structures that the three aforementioned sources of interdependence manifest in the studies of democratization. More specifically, first, I develop a multivariate event history (duration) model to test for the existence and the strength of the dependency between the emergence and breakdown of democracies. Next, I develop a new spatial econometric model to test for the existence of the conventional contagion processes and the reinforcement dynamic in democratization. The ability to distinguish statistically the conventional contagion dynamic and the selection/reinforcement dynamic is one of the major contributions of this new spatial model.

1.1.5 Plan

Essay 1 "Interdependent Duration Models in Political Science: An Application to the Democratic Transition and Survival in Africa" introduces the multivariate duration model that I developed in Hays and Kachi (2011). I apply the method to the study of democratic transitions and consolidation. More broadly, the multivariate duration model is useful when the timing of multiple political events or a single political event occurring to multiple units is interdependent across events or units. Specifically, in my application, the particular events of interest are the emergence and collapse of democracies. In a way consistent with earlier democratization studies (e.g., Przeworski et al. 2000), I recognize that each of these events is explained by the event- and unit-specific attributes, such as economic and political-institutional variables for each event and country-year. However, as explained earlier, a game-theoretic literature of democratization (e.g., Acemoglu and Robinson 2006*a*) implies a causal relationship from the (potential, anticipated) survival of democracies to the timing of democratization. At the same time, scholars who advocate the importance of gradual transitions to democracy emphasize the negative effect of democratic transitions under premature socio-economic conditions on the stability of future democracies.

They imply a possible causal effect from the timing of democratization to the survival of democracies. The potential two-way causality implied by these theoretical works is the reason why we need to utilize a systems-of-equations approach, instead of a single-equation approach (for each observation unit).

Since the dependent variables of interest are about timing (duration, or the rate of failure of a type of regime), this leads me to develop a multivariate duration model, in which one can estimate the effects of country-specific variables on the duration dependent variables, by, at the same time, statistically testing for the presence and evaluating the strength of the across-duration dependency.

Essay 2 "Right-Censoring in Interdependent Duration Models with the Systems of Duration Equation Modeling (SDEQ) Approach" extends the interdependent duration model developed in Essay 1 to account for *right-censoring*. To clarify the terminology, we say that the observation is censored when an observation's full survival history is not observed, due to, for example, the time frame of the data collection. For example, if one is interested in the survival of democracies, the continuing democracies at the end of the data-collection date are right-censored (e.g., Alemán and Yang 2011; Svolik 2008). If we look at the duration of cabinet survival, there are two sources of right-censoring: first, cabinets that are still surviving at the end of the collected data are right-censored observations, and second, cabinets that lasted until the end of the constitutional inter-election period (CIEP) and "had to end" are also right-censored, because they could have lasted longer but we do not have complete information about their survival durations due to the imposed cut-off (e.g., Laver and Shepsle 1996; Lupia and Strøm 1995; Warwick and Easton 1992; Alt and King 1994; Diermeier and Stevenson 1999). Generally, the existence of right-censored observations can be, in fact, one of the reasons to use the duration approach, because a modification in likelihood functions in duration models can account for censoring (Box-Steffensmeier and Jones 2004).

The idea of this treatment for right-censored observations is fairly simple to conceptualize and implement in single-duration models and interdependent duration models that are based on copulas. It is, on the contrary, not mathematically straightforward to account for right-censoring in the SDEQ approach. However, remember that the SDEQ approach has an advantage of modeling interdependence in a substantively meaningful way, compared to the copula approach. Therefore in this paper, I make a first attempt to develop a relatively simple way to evaluate an interdependent-duration likelihood function with a treatment that can account for right-censoring. Given the ubiquity of duration interdependence (see Essay 1) and right-censored observations in political science (Box-Steffensmeier and Jones 2004), the likelihood I derive here should be broadly applicable.

Just like in Essay 1, the key mathematical strategy to write the likelihood function for right-censored observations turns out to be the change of variables theorem, this time, however, on multivariate integrals. In short, the multivariate integrals (over the number of dimensions equivalent to the number of duration processes in the model) are the sources of difficulty in writing this likelihood function.

It is broadly believed that one cannot eliminate the integrals analytically, and hence estimation needs to be done by simulations. This belief seems to stem from the fact that the likelihoods for multivariate or spatial-lag probit models, for example, contain multiple integrals and they indeed require simulation approaches to estimate (e.g., **?**Hays 2009). However, this is only due to the fact that the probit likelihood function contains normal cumulative distribution functions (CDF). Normal CDF contains an error function ("erf"), which in turn contains a integral, regardless of the dimensionality. If the error terms of the structural form have "simpler" distribution than normal, then it is possible to eliminate multiple integrals in the likelihood using the error i.i.d. assumption and the change of variables theorem.

[Edit later] I will also note the applicability and generalizability of this mathematical transformation to derive analytically the likelihood function for multivariate logit and spatial-lag logit models. It is also useful to develop an interdependent "duration" model, where some of the dependent variables are durations but others are binary; for example, one might want to address the dependency between the introduction of multiparty elections (by dictators or democratic leaders) and the survival of such democratic institutions. Since binary outcomes are common in political science, these estimators would be extremely useful for applied researchers.

Essay 3 "A Spatial Econometric Approach to Coevolution: The Diffusion and Reinforcement of Political Regimes" introduces a new spatial estimator, an MSTAR+C model, developed in Hays, Kachi and Franzese (2010) and applies it to the study of the diffusion and reinforcement of democracies. This model deals with the type of interdependence that arises when units are connected through various kinds of networks, and political behavior of these units are contagious through the networks. For example, if citizens in these countries learn about the odds of successful revolution/riots against the authoritarian government by observing similar incidences in other countries, revolution can spread. Moreover, it is likely that countries that are more strongly dependent influence each other's politics more significantly than countries that are less dependent on each other, as Tobler's Law summarizes very succinctly: "everything is related to everything else, but near things are more related than distant things." Note that the distance here does not have to be geographical. Countries that are geographically far away from each other could have stronger economic ties (and therefore stronger political influences). The notion of explicit interdependence is the key of diffusion studies. The first merit of taking a spatial econometric approach is that we do not lose such "relational" dynamics specified in the theoretical model, unlike most statistical approaches in existing studies of "diffusion".

Furthermore, my account departs from the existing concept of diffusion by allowing for the possibility that countries form implicit regime-support ties, reinforcing each others' current regime types (selection). In this mechanism, countries' regimes affect with whom they interact in the future. Such influence that occurs on the self-selective network generates a particular dynamic over time. In this essay, I refer to it as the *network-behavior coevolution*. The networks (of similar regimes) and countries' behavior (their regime choice) recursively feed in each other and evolve together over time. This becomes the source of regime reinforcement and potentially the source of the democracy and autocracy clusters in the world. The second, and the most important contribution of the present spatial model is the ability to distinguish this reinforcement mechanism from the conventional (and simpler) contagion mechanism, and estimate the strength of each.

Chapter 2

Essay1
Interdependent Duration Models in Political Science: An Application to the Democratic Transition and Survival in Africa

Abstract

Interdependent duration processes are common in politics and other strategic settings. The time to one type of political event frequently depends on the time to another related event, and the time to an event for one actor often depends on the time to that same event for others. Put in a slightly different way, politics and strategic behavior generate interdependence across durations and duration interdependence across actors. In order to test for the existence and the strength of the dependence between multiple durations as well as the effects of covariates explaining each duration, I develop a generalized parametric simultaneous equations model, and derive the corresponding full information maximum likelihood (FIML) estimator based on the Weibull distribution. In this essay, I refer to the new model as an "interdependent duration model." I show with Monte Carlo experiments that the new interdependent duration estimator outperforms the alternatives available to those doing applied empirical research. I compare the following three traditional approaches to the new model. Naive estimators that either ignore the interdependence among duration processes or treat one as exogenous to the others are badly biased when the true relationships are simultaneous ones. Two-stage least squares, while consistent, is highly inefficient relative to the FIML.

I illustrate these findings in a study of the determinants of democratic transitions and consolidation in Africa. I first present three distinct theories that suggest the interdependence between the liberalization duration and the survival of a democracy that comes after the liberalization period. The first mechanism, where the authoritarian elites more willingly liberalize the regime when they anticipate a short-lived democracy in the future, suggests a

positive causal effect of the survival on liberalization. The second mechanism is such that the democratic opposition more actively and willingly commits to costly political actions such as revolutions and the demand for new constitutions, when they anticipate a successful transition to a long-lived democracy. This implies a negative causal effect of the survival on the liberalization duration. The third mechanism is where a longer liberalization duration helps the society develop cultures that appreciate democratic values and hence leads to a long-surviving democracy. This implies a positive causal effect of the liberalization duration on the survival of democracies. The empirical findings suggest that the duration of liberalization and the survival of democracies are interdependent, and the causal relationships between the two durations exist in both directions. From the sign of the causal relationships, it is indicative that the democratic opposition does form anticipation about the future prospect of its anti-government movements, affecting the duration of the liberalization period, in a way the potential of a more durable democracy in the future shortens the duration of the democratic transition. At the same time, a longer transition duration positively influences the survival of a future democracy. The qualitative difference between then empirical findings from my new approach and the traditional approach is stark, suggesting that employing traditional empirical approaches could lead one to the wrong substantive conclusion.

2.1 Overview

In many topics in political science, understanding survival time and failure rates of certain political processes is crucial. To name just a few, comparative schoalars have examined the survival and dissolution of cabinets in parliamentary democracies (King et al. 1990; Warwick 1992), the duration of political regimes (Chapman and Roeder 2007; Svolik 2008; Alemán and Yang 2011), and the timing of union-friendly labor reforms (Murillo and Schrank 2005). In international relations, scholars have examined the survival of military alliances (Bennett 1999), post-conflict peace duration (Fortna 2004; Werner and Yuen 2005), and the speed at which policies diffuse around the world (Simmons and Elkins 2004). Examples from American politics include the time until major pieces of legislation are amended (Maltzman and Shipan 2008), the duration of Supreme Court nominations (Shipan and Shannon 2003), and the timing of issue position taking in Congress (Box-Steffensmeier, Arnold and Zorn 1997; Boehmke 2006; Darmofal 2009). In many of these situations, we are interested in, on average, how long a certain political phenomenon will survive *given that it has reached the current time period*. Or equivalently, we are interested in how likely it is that the political event of our interest fails in the current time period *given that it has survived until now*. This time dependence requires a special methodological care. Event history analysis offers a set of powerful tools that allow us to assess whether a certain political process is at risk of experiencing some failure–or more generally "event"–at a given time point.

Furthermore, (a) the occurrence of multiple political events or (b) a single political event occurring to multiple units might be interdependent across events or units. For example, imagine a situation where states bargain over an international cooperation agreement, just as described in Fearon (1998). Intuitively, the survival of such agreements (i.e., the effectiveness of the enforcement scheme) could depend on the length of the bargaining phase: for example, cooperation can be more successful (longer survival of agreements) because parties negotiate over the terms carefully (longer bargaining), or agreements might fail sooner (shorter survival of agreements) reflecting the parties' divergent preferences over the terms (longer bargaining). Dependency of the two durations, the bargaining and enforcement phases, can go the other way around as well. As Fearon (1998) describes, states can be more willing to hold out in the bargaining phase negotiating harder when they optimistically anticipate the survival of cooperation. In this case, the expected longer survival of an agreement induces a longer bargaining process. This example belongs to case (a), in which the length of two distinct duration processes (bargaining and enforcement) are dependent on each other. Similarly, the timing of democratization might be a significant predictor of the survival of newly-emerged democracy. Prolonged durations of regime transition periods could imply greater challenges to democratization and hence fragile democracies. Or on the contrary, slower transitions might lead to more stable democracies. At the same time, if the political actors (elites or the masses) have certain expectations about the survival of future democracies, the expectations can affect the speed and intensity of transition periods. For researchers (econometricians), these anticipations are unobservable, but they affect the data generating process regardless. This potential interdependence in democratic transitions and consolidation is the study that I pursue in the application section.

A good example for case (b), where a single event occurring to multiple observational units, is the timing of issue position taking in Congress, the event of our interest is a single piece of legislation. However the timing of decision making by multiple Congress members (i.e., multiple observational units) can be interdependent. This dependency can, for example, stem from "unobserved factors such as Presidential lobbying and/or party loyalty influence" that make Congress members' decision making (both the timing and contents) interdependent behind the scene (Boehmke 2006).

In either case, whether (a) or (b), once we attempt to take into account the interdependence across multiple duration processes, a recursive influence dynamic emerge as depicted in Figure 2.1. First, there is a whole system of multiple duration processes. For simplicity, I use an example of a system of only two duration processes, y's. These multiple durations can be the length of regime transition period and the survival of democracies, or the spells of time taken by two Congress members till they announce their issue positions. Typically, scholars are interested in testing for the effects of structural predictors X's on either of the multiple duration processes y's. For instance, a question of whether economic development has any influence on the timing of democratic transitions or on the survival of democracies belongs to this type of empirical question.

However, once we start theorizing the dependency among durations (e.g., the transition duration explaining the survival and vice versa), the right-hand side of each duration equation

Figure 2.1: A common research framework with interdependent duration processes

$$\begin{cases} \text{Duration 1: } \mathbf{y}_1 \leftarrow \mathbf{X}_C + \mathbf{X}_1 + \text{Duration 2: } \mathbf{y}_2 + \boldsymbol{\varepsilon}_1 \\ \\ \text{Duration 2: } \mathbf{y}_2 \leftarrow \mathbf{X}_C + \mathbf{X}_2 + \text{Duration 1: } \mathbf{y}_1 + \boldsymbol{\varepsilon}_2 \end{cases}$$

contains the duration processes that other equations predict (as dependent variables), generating a endogenous influence loop between the two equations. This feedback loop between the two processes makes it impossible for us to estimate any of the included parameters without biases, if we analyze the two processes separately.

Unfortunately, to this point, most studies have treated multiple duration processes in isolation, meaning that they study either of the two durations in Figure 2.1, but not both. At worst, such single equation studies suffer from simultaneity biases, if the "true" model is with the across-duration dependency. Simply put, the bias comes from ignoring the feedback effects between the two endogenous duration variables. When a shock (some change, whether an increase or decrease) is given to y_2 on the right-hand side of the first equation, for example, it affects the left-hand side y_1. In turn, this change in y_1 on the right-had side of the second equation affects the y_2 of the left-hand side. If the system is convergent, this cycle will stop at the long-run equilibrium of the system (dynamic equilibrium). The system can also be explosive. Either way, it is possible to take into account this feedback effect, only if we study the whole system in a unified model. This is why it is important to develop a multivariate duration model, given the aforementioned substantive arguments that suggest the dependency between the two durations.

This essay consists of two parts. First, Section 2.2 through 2.5 are devoted to the development of a new duration estimator. The second part, Section 2.6, is devoted to the application of the estimator. In the first part, I review some of the existing empirical strategies for modeling interdependent durations, highlighting the important differences between simultaneous equations (SEQ) and seemingly unrelated regression (SUR) models (or substantive and nuisance models of interdependence more generally) as well as the relative strengths and weaknesses of using change-of-variables-based and copula-based likelihoods. Second, I present a general simultaneous equations model of interdependent durations and derive the corresponding FIML estimator. This model encompasses both traditional SEQ models and (spatial) duration models. Third, I compare the performance of several estimators against the new FIML using Monte Carlo experiments. In the second part, I first motivate the study of democratization and consolidation in countries in Africa between 1956 and 2001, emphasizing the potential dependency between the two processes. Next I describe the variables and data, and present the estimation results. Finally, I compare the new interdependent duration approach with the traditional univariate approach to highlight the substantive (or qualitative) difference in conclusions that we would make based off of the two approaches. I conclude by summarizing both the methodological and substantive contributions of this essay. The appendices for Essay 1 include detailed mathematical derivations of quantities used in the methodology section (Appendix A.1.1- A.1.2), a summary of other applications that I conducted for a working paper (Hays and Kachi 2011) (Appendix A.1.4), and the Stata code used to estimate the models for the application in this essay (Appendix A.1.3).

2.2 Existing Strategies for Modeling Interdependent Durations

How should I model interdependent durations? There are two basic approaches to modeling (the two kinds of) interdependence. One approach is to assume that the interdependence arises in the stochastic part of the model only. The second posits full interdependence in both the stochastic and systematic components. Often, strategies of the first variety are called nuisance approaches, while those of the second are referred to as substantive models of interdependence.[1] Together with the two kinds of interdependent duration processes–

[1] Occasionally, these labels are inaccurate. For example, many scholars treat unobservables as substantively important (see Boehmke 2006). The interdependence is in the disturbances, but it is central to the

interdependence across durations and duration interdependence across actors–this gives a four-fold model typology. Examples include SUR, spatial error, SEQ, and spatial lag models respectively, and each of these can be found in the literature.

Quiroz Flores (2008), for instance, uses copula functions to estimate a SUR model of the tenure of chief executive officers and the median tenure of their ministers. The argument is that there are unobservable common shocks that affect the tenure of both chief executives and their ministers. Flores finds significant correlation in tenures and shows with Monte Carlos that an estimator that accounts for cross-equation correlation in disturbances is more efficient than those that do not. Darmofal (2009) estimates a spatial duration model for issue position taking on NAFTA in the U.S. Congress. This work, which I discuss in more detail below, is representative of the spatial-error approach to interdependence. His model allows for individual and shared frailties. Specifically, it is a model of spatially autocorrelated random effects. Darmofal finds strong evidence for state-level shared frailties in the timing of issue position taking.

An early example of interdependence across durations is found in Lillard (1993). He presents a simultaneous equations model of marriage duration and fertility timing. In his setup, the hazard of marriage dissolution has a direct effect on the fertility hazard, and prior outcomes of the fertility process affect the dissolution hazard. The baseline hazards are a function of piecewise linear splines, which allows for a more flexible form. Honoré and de Paula (2008) provide a recent example of duration interdependence across actors. They derive an interdependent duration model from a strategic two-player game. In their model, agents choose how long to participate in a particular activity before switching to an alternative activity. The utility from switching for one agent depends on whether the other agent has switched. There are many examples of this form of interdependence in microeconomics including the adoption of new technologies and market entry decisions by firms. With Monte Carlo experiments, they show, among other things, that treating endogenous durations as

analysis nonetheless.

exogenous typically leads to an overestimation of strength of interdependence.

From my perspective, the main problem with nuisance approaches to duration dependence is that they fail to capture the forms of strategic interdependence that we typically have in mind: that duration i (or the duration for unit i) is a function of duration j (or the duration for unit j). The problem with the substance approaches is that they can lead to models that are more difficult to estimate. In other words, we see a potential trade-off between the conceptual match of the empirical models with theory and the ease with which these models can be estimated. Given this trade-off, I develop the simplest possible SEQ FIML estimator.

In answering the question about how to model duration interdependence, one must keep in mind the connection between structure and substance, on the one hand, and structure and estimation, on the other. I discuss these connections over the next several sections, beginning with structural assumptions about outcome dependency. With respect to estimation, I show how the SUR structure leads naturally to copula-based estimators while the SEQ structure makes the change-of-variables approach, because of its relative simplicity, attractive to those doing applied research.

2.2.1 Assumptions about Dependency Structure: SUR vs. SEQ

SUR

The seemingly unrelated regressions (SUR) and simultaneous equations (SEQ) models make very different assumptions about the structure of outcome dependency. Understanding these differences is crucial to choosing the right model and estimator.

First, the association among the outcome variables, $\mathbf{y}$'s, can be driven solely by their stochastic components that are generated from a single joint probability distribution. The SUR model captures this type of dependency among outcomes. In matrix notation, the SUR model takes the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{y}$ is an $ND \times 1$ vector containing N observations on D endogenous variables. The matrix $\mathbf{X}$ contains $ND \times K$ observations on K exogenous variables, where $K = \sum_{d=1}^{D} k_d$, k_d being the number of exogenous variables in the equation for the d^{th} endogenous variable, and $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients on them. The final term $\boldsymbol{\varepsilon}$ in equation (2.1) is an $ND \times 1$ vector of disturbances with covariance matrix $\mathbf{V}_{SUR} = \mathbf{\Sigma I}$. Rewriting (2.1) in terms of its constituent equations, we have

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_D \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_D \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_D \end{bmatrix}, \tag{2.2}$$

where $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_D]'$ is generated from a single joint distribution with the covariance matrix

$$\mathbf{V}_{SUR}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1D} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{D1} & \sigma_{D2} & \cdots & \sigma_{DD} \end{bmatrix} \mathbf{I}. \tag{2.3}$$

Due to the jointly-generated disturbances, outcomes y's seem to be related to each other. It should be noted, however, that the dependency structure implied by this model is different from what we have in mind when we say y_i depends on y_j $(i \neq j)$ –there is no component in this model that captures the relationship $y_i = f(y_j)$.

SEQ

On the contrary, the SEQ approach models the explicit dependency among outcomes. It takes the form

$$\mathbf{y} = \mathbf{AI}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.4}$$

which written in terms of its constituent equations is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \alpha_{12}\mathbf{I} & \cdots & \alpha_{1D}\mathbf{I} \\ \alpha_{21}\mathbf{I} & \mathbf{0} & \cdots & \alpha_{2D}\mathbf{I} \\ & & \vdots & \\ \alpha_{D1}\mathbf{I} & \alpha_{D2}\mathbf{I} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_D \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_D \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_D \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_D \end{bmatrix}. \tag{2.5}$$

The $\mathbf{AI}_{(ND\times ND)}$ matrix that consists of $\alpha_{ij}\mathbf{I}_{(N\times N)}$ $(i \neq j)$ represents the degree of direct dependency among $\mathbf{y}_i$'s. For example, $\alpha_{ij}\mathbf{I}_{(N\times N)}$ denotes the effect of $\mathbf{y}_j$ on $\mathbf{y}_i$.

Since the endogenous variable $\mathbf{y}$ is now on the right-hand side of the structural equations (2.5), one needs to derive the reduced form in order to discuss properties of the stochastic component. Equation (2.5) can be written in reduced form as

$$\begin{aligned} \mathbf{y} &= (\mathbf{I} - \mathbf{AI})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{AI})^{-1}\boldsymbol{\varepsilon} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \end{aligned} \tag{2.6}$$

where $\boldsymbol{\Gamma} = (\mathbf{I} - \mathbf{AI})^{-1}$ and the covariance matrix for $\mathbf{u}$, the vector of reduced form disturbances, is

$$\mathbf{V}_{SEQ}(\mathbf{u}) = \boldsymbol{\Gamma}'\mathbf{V}\boldsymbol{\Gamma}. \tag{2.7}$$

An important difference between the SUR and SEQ models is that the covariances in (2.7) are a function of the coefficients in $\mathbf{A}$. In other words, the variances and covariances among the reduced form disturbances have to be consistent with the structural relationships among the endogenous variables. This feature has implications for estimating the two models.

In the next two sections, I construct maximum likelihood (ML) estimators based on the SUR and the SEQ assumption respectively. For the SUR approach, I use copulas to obtain the necessary joint probability densities for the likelihood function, and for the SEQ approach, I employ the change-of-variables technique to derive the necessary joint densities. These two methods look unrelated at first sight. However, the resulting likelihood functions are comparable and they exemplify the different assumptions about the dependency structure that I make by choosing either the SUR or the SEQ framework. I focus on this comparison below in section 1.4.

2.2.2 Copula-Based Likelihood for a Duration SUR Model

A copula is a function that gives a proper joint distribution function from univariate marginal distribution functions. Several papers in political science use copulas or copula related distributions to derive likelihoods for empirical analysis including Boehmke, Morey and Shannon (2006), Boehmke (2006), Quiroz Flores (2008) and Fukumoto (2009) among others. The primary advantage of using copulas is that one has the joint distribution function, which is necessary to construct many likelihoods–e.g., the likelihoods for qualitative or limited dependent variables models. The main disadvantage of using copula-based distributions is that the covariance structures are constrained. These constraints imply limits on the range of association among the variables, and they also make it difficult to use copulas to derive likelihoods for SEQ models.

First, consider a joint distribution function of random variables y_1^* and y_2^* generated from the following Farlie-Gumbel-Morgenstern (FGM) copula

$$F(y_1^*, y_2^*) = F(y_1^*)F(y_2^*)[1 + \alpha\{1 - F(y_1^*)\}\{1 - F(y_2^*)\}], \tag{2.8}$$

where α, the association parameter, captures the degree of dependence between the two y^*'s and $-1 \leq \alpha \leq 1$. The corresponding joint density function is given as

$$f(y_1^*, y_2^*) = f(y_1^*)f(y_2^*)[1 + \alpha\{2F(y_1^*) - 1\}\{2F(y_2^*) - 1\}]. \tag{2.9}$$

For example, given the univariate Weibull distribution function and density function

$$\begin{aligned} F(y^*) &= 1 - e^{-(\frac{y^*}{\theta})\lambda} \\ f(y^*) &= \frac{\lambda}{\theta}\Big(\frac{y^*}{\theta}\Big)^{\lambda-1} e^{-(\frac{y^*}{\theta})^\lambda}, \end{aligned} \tag{2.10}$$

the copula (2.8) and (2.9) generates the joint cumulative and the joint density functions of the bivariate Weibull distribution.

$$\begin{aligned} F(y_1^*, y_2^*) &= (1 - e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1}})(1 - e^{-(\frac{y_2^*}{\theta_2})^{\lambda_2}})(1 + \alpha e^{-(\frac{y_1^*}{\theta_1})^{\lambda_1} - (\frac{y_2^*}{\theta_2})^{\lambda_2}}) \\ f(y_1^*, y_2^*) &= \frac{\lambda_1}{\theta_1}\frac{\lambda_2}{\theta_2}\Big(\frac{y_1^*}{\theta_1}\Big)^{\lambda_1-1}\Big(\frac{y_2^*}{\theta_2}\Big)^{\lambda_2-1} e^{-2[(\frac{y_1^*}{\theta_1})^{\lambda_1} + (\frac{y_2^*}{\theta_2})^{\lambda_2}]} \\ &\quad \times [4\alpha - 2\alpha e^{(\frac{y_1^*}{\theta_1})^{\lambda_1}} - 2\alpha e^{(\frac{y_2^*}{\theta_2})^{\lambda_2}} + (1+\alpha)e^{(\frac{y_1^*}{\theta_1})^{\lambda_1} + (\frac{y_2^*}{\theta_2})^{\lambda_2}}], \end{aligned} \tag{2.11}$$

where $y_1^* \geq 0$, $y_2^* \geq 0$, $-1 \leq \alpha \leq 1$, $\theta_1 > 0$, $\theta_2 > 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$. Again, α is a dependence parameter, which induces the correlation between y_1 and y_2, and the λ's are shape parameters that determine the curvature of the distribution. θ's are scale parameters. Note that this becomes the second of the third Gumbel (bivariate exponential) distribution discussed in Gumbel (1960) when $\theta_1 = \theta_2 = 1$ and $\lambda_1 = \lambda_2 = 1$.

With the FGM bivariate Weibull distribution, the degree of admissible linear association between the variables, Pearson's correlation, is limited to $-0.322409 \leq \rho \leq 0.322409$. I derive the possible range for ρ and summarize some of the mathematical properties of this bivariate Weibull distribution in Appendix A.1.1. This constraint may or may not be a serious limitation depending on the true strength of interdependence among the durations one is modeling. A second, more serious concern stems from the difficulty in transforming the copula-based SUR estimator into a SEQ estimator. To see this, we need to compare SUR and SEQ likelihoods.[2] Below is the likelihood function for a SUR model as described above, where there are two duration processes ($D = 2$) with N observations for each duration;

$$
\begin{aligned}
L(\mathbf{X}, \boldsymbol{\beta}, \lambda_1, \lambda_2 | y_1^*, y_2^*) &= \prod_{i=1}^{N} f(y_{i1}^*, y_{i2}^*) \\
&= \prod_{i=1}^{N} \Big(\frac{\lambda_1}{\theta_1} \frac{\lambda_2}{\theta_2} (\frac{y_{i1}^*}{\theta_1})^{\lambda_1 - 1} (\frac{y_{i2}^*}{\theta_2})^{\lambda_2 - 1} e^{-2[(\frac{y_{i1}^*}{\theta_1})^{\lambda_1} + (\frac{y_{i2}^*}{\theta_2})^{\lambda_2}]} \\
&\times [4\alpha - 2\alpha e^{(\frac{y_{i1}^*}{\theta_1})^{\lambda_1}} - 2\alpha e^{(\frac{y_{i2}^*}{\theta_2})^{\lambda_2}} + (1+\alpha) e^{(\frac{y_{i1}^*}{\theta_1})^{\lambda_1} + (\frac{y_{i2}^*}{\theta_2})^{\lambda_2}}] \Big),
\end{aligned}
\tag{2.12}
$$

where the θ_d's, the scale parameters, are equal to $e^{\mathbf{X}_d \boldsymbol{\beta_d}}$ and the λ_d's are the shape parameters.

2.3 New Estimator: Change-of-Variable Likelihood for an SEQ Weibull Duration Model

This section present a general simultaneous equations model for interdependent duration processes and derive its full information maximum likelihood estimator. I then return to the SUR likelihood for purposes of comparison in section 1.4.

[2]ML estimation using this joint distribution offers a possible solution to the problems caused by unobservables (mainly inefficiencies) from which the existing literature may suffer. Unobservables are much more pernicious in the selection model context because they are a potential source of endogeneity and bias (see Boehmke, Morey and Shannon 2006). Copulas like the FGM are helpful here because the selection bias correction does not require covariance decomposition.

Linear Parameterization of Weibull Durations (The AFT Model)

The dependent variables of interest, $\mathbf{y}^*$, are D distinct duration processes that have Weibull distributions with two parameters.

$$y^*_{id} \sim Weibull(\lambda_d, \theta_d), \tag{2.13}$$

where $i = \{1, \cdots, N\}$ denotes the observational-unit index and $d = \{1, \cdots, D\}$ denotes the duration index, implying that there are $N \times D$ observations in total. The notation λ is the shape parameter and the θ is the scale parameter. These distributional parameters take common values across N observational units; hence they have only one subscript that indicates duration process. A common way to parameterize a Weibull model of D interdependent durations is to log-linearize the model and obtain a log-linear system of D equations (Box-Steffensmeier and Jones 2004, p.26). It is also known that the logged Weibull variable turns out a standard Gumbel variable that is scaled by the shape parameter in the original Weibull distribution.[3] For example, in the univariate Weibull case, the log-linear form would look like

$$\begin{aligned} y = \ln y^* &= \ln\theta + \frac{1}{\lambda}\varepsilon \\ &= \mathbf{X}\boldsymbol{\beta} + \frac{1}{\lambda}\varepsilon, \end{aligned} \tag{2.14}$$

[3] The standard Gumbel distribution is a special case of the type-I extreme value (minimum) distribution. The distribution and density functions of the type-I extreme value (minimum) distribution are

$$\begin{cases} f(u) = \frac{1}{b} e^{\frac{u-a}{b}} e^{-e^{\frac{u-a}{b}}} \\ F(u) = 1 - e^{-e^{\frac{u}{b}}}, \end{cases}$$

where a is the location parameter and b is the scale parameter. The distribution and density functions of the standard Gumbel distribution are

$$\begin{cases} f(u) = e^{u} e^{-e^{u}} \\ F(u) = 1 - e^{-e^{u}}. \end{cases}$$

Note that the standard Gumbel distribution is a special case of the type-I extreme value distribution, where $a = 0$ and $b = 1$. A logged Weibull variable has the type-I extreme value distribution in general and only the scaling of the resulting extreme value variable varies depending on how one sets the scale parameter of the extreme value variable. For further details, see Appendix 2.

where $\varepsilon \sim ExtremeValueI\ (StandardGumbel)$ and I define $y = \ln y^*$. The second line of equation (3.3) shows how one could include covariates, by making the Weibull scale parameter, θ, a function of the covariates, $\theta = e^{\mathbf{X}\boldsymbol{\beta}}$. For further detail regarding the link between a Weibull and an extreme value distribution, see Appendix 2.

The System

The system of D distinct durations with N observational units in matrix notation is

$$\mathbf{y}_{(ND\times 1)} = \mathbf{A}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}. \tag{2.15}$$

The dependent variable, $y_{id} = \ln \mathbf{y}^*_{id}$, is a logged Weibull random variable. The vector $\mathbf{y}$ is a stack of D vectors, each of which contains N observational units.

$$\mathbf{y}_{(ND\times 1)} = \begin{pmatrix} \mathbf{y}_{.1} \\ \vdots \\ \mathbf{y}_{.D} \end{pmatrix}, \text{ where } \mathbf{y}_{.d(N\times 1)} = \begin{pmatrix} y_{1d} \\ \vdots \\ y_{Nd} \end{pmatrix}.$$

The matrix $\mathbf{A}$ is the coefficient matrix for the dependence. An element matrix $\boldsymbol{\alpha}^{d'}_{.d}$ contains coefficients representing the effects of the duration d on the duration d'. The diagonal elements $\boldsymbol{Sp}$'s in the $\mathbf{A}$ matrix are the matrices that capture the "spatial" dependency. This is the dependency among N observational units within each duration process. I call it "spatial" dependency for convenience, because the linear system captures the among-unit dependency using weights matrices just like in spatial contexts. Note that $\boldsymbol{Sp}^d = \mathbf{0}$ for all d when one assumes no among-unit dependency. Similarly $\boldsymbol{\alpha}^{d'}_{.d} = \mathbf{0}$ when one assumes no dependency between duration d and d'.

$$\mathbf{A}_{(ND\times ND)} = \begin{pmatrix} \boldsymbol{Sp}^1 & \boldsymbol{\alpha}^1_{.2} & \cdots & \boldsymbol{\alpha}^1_{.D} \\ \boldsymbol{\alpha}^2_{.1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^D_{.1} & \cdots & \boldsymbol{\alpha}^D_{.D-1} & \boldsymbol{Sp}^D \end{pmatrix},$$

where

$$\boldsymbol{\alpha}^{d'}_{.d(N\times N)} = \begin{pmatrix} \alpha^{d'}_{.d} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \alpha^{d'}_{.d} \end{pmatrix}, \boldsymbol{Sp}^d_{(N\times N)} = \begin{pmatrix} 0 & \alpha^d_{(1,2)} & \cdots & \alpha^d_{(1,N)} \\ \alpha^d_{(2,1)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha^d_{(N-1,N)} \\ \alpha^d_{(N,1)} & \cdots & \alpha^d_{(N,N-1)} & 0 \end{pmatrix}$$

The vector $\mathbf{x}$ denotes a set of covariates and the supscript indicates to which equation the covariate vector is specific. Each vector $\mathbf{x}$ contains K covariates with coefficients denoted by β. The subscript of $\mathbf{X}$, $.d$, indicates that these x's affect duration d, and the number of covariates, i.e., the number of variables in each $\mathbf{X}_{.d}$ is denoted K_d. For example, the duration 1 is partially predicted by a set of covariates $X_{.1}$, and there are K_1 variables included. The error term u_{it} in this structural form is i.i.d. with the extreme value minimum distribution. The error term is multiplied by $\lambda^{-1}_{.d}$, which is the shape parameter of the original Weibull distribution and the value of λ is allowed to vary across duration processes.

$$\mathbf{X}_{(ND\times(K_0+\cdots+K_D))} = \begin{pmatrix} \mathbf{X}_{.1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{.2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}_{.TD} \end{pmatrix}, \text{ where } \mathbf{X}_{.d(N\times K_d)} = \begin{pmatrix} x^1_{1d} & \cdots & x^{K_d}_{1d} \\ \vdots & \ddots & \vdots \\ x^1_{Nd} & \cdots & x^{K_d}_{Nd} \end{pmatrix}$$

$$\boldsymbol{\beta}_{(1+K_1+\cdots+K_D\times 1)} = \left(\begin{array}{c|ccc|ccc|ccc|ccc} \beta_{.0} & \beta_{.1}^1 & \cdots & \beta_{.1}^{K_1} & \beta_{.2}^1 & \cdots & \beta_{.2}^{K_2} & \cdots & & \cdots & \beta_{.D}^1 & \cdots & \beta_{.D}^{K_D} \end{array} \right)',$$

where $\beta_{.0}$ is a constant;

$$\mathbf{L}_{(ND\times ND)} = \begin{pmatrix} \mathbf{L}_{.1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{L}_{.D} \end{pmatrix}, \text{ where } \mathbf{L}_{.d(N\times N)} = \begin{pmatrix} \frac{1}{\lambda_{.d}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{\lambda_{.d}} \end{pmatrix};$$

$$\mathbf{u}_{(ND\times 1)} = \begin{pmatrix} u_{11} \\ \vdots \\ u_{ND} \end{pmatrix}.$$

The following reduced form can be derived from the structural form (3.5);

$$\begin{aligned} \mathbf{y}_{(ND\times 1)} &= (\mathbf{I}-\mathbf{A})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}-\mathbf{A})^{-1}\mathbf{L}\mathbf{u} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{L}\mathbf{u} \\ &= \boldsymbol{\Gamma}\mathbf{X}\boldsymbol{\beta} + \mathbf{v}, \end{aligned} \tag{2.16}$$

where $\boldsymbol{\Gamma} = (\mathbf{I}-\mathbf{A})^{-1}$ and $\mathbf{v} = \boldsymbol{\Gamma}\mathbf{L}\mathbf{u}$.

Deriving the Likelihood via Change of Variables

The only task left before writing a likelihood function is to derive the joint density of y's. We do not know the joint distribution of y's, but fortunately it is not hard to obtain the joint distribution of u's, because they are assumed to be i.i.d and we know that the marginal of u has the type I extreme value distribution. I use the change of variables theorem to derive the joint pdf of y's from the joint pdf of u's. By solving equation (2.16) for $\mathbf{u}$, we have

$$\mathbf{u}_{ND\times 1} = g^{-1}(\mathbf{y}) = (\mathbf{\Gamma L})^{-1}\mathbf{y} - \mathbf{L}^{-1}\mathbf{X}\boldsymbol{\beta}. \tag{2.17}$$

The Jacobian matrix of $g^{-1}(\mathbf{y})$ is

$$\mathbf{J} = \begin{pmatrix} \frac{\partial g_{11}^{-1}(\mathbf{y})}{\partial y_{11}} & \cdots & \frac{\partial g_{11}^{-1}(\mathbf{y})}{\partial y_{ND}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{ND}^{-1}(\mathbf{y})}{\partial y_{11}} & \cdots & \frac{\partial g_{ND}^{-1}(\mathbf{y})}{\partial y_{ND}} \end{pmatrix}.$$

If the inverse vector function, $(u_{11}, \cdots, u_{ND}) = g^{-1}(y_{11}, \cdots, y_{ND})$, exists for all $\mathbf{y} = (y_{11}, \cdots, y_{ND})$ such that $\mathbf{y} \in \{\mathbf{y} = g(\mathbf{u})\}$, the joint density of $\mathbf{Y} = g(\mathbf{U})$ is given by

$$\begin{aligned} h(y_{11}, \cdots, y_{ND}) &= \begin{cases} f(g_{11}^{-1}(y_{11}, \cdots, y_{ND}), \cdots, g_{ND}^{-1}(y_{11}, \cdots, y_{ND}))|det(\mathbf{J})| \\ 0, \text{ otherwise} \end{cases} \\ &= \begin{cases} f(u_{11}, \cdots, u_{ND})|det(\mathbf{J})| \\ 0, \text{ otherwise} \end{cases} \\ &= \begin{cases} f(u_{11})f(u_{12})\cdots f(u_{ND})|det(\mathbf{J})| \\ 0, \text{ otherwise.} \end{cases} \end{aligned} \tag{2.18}$$

The last line in equation (2.18) follows from the i.i.d. assumption of u, and each $f(u_{id})$ is the standard Gumbel pdf.

The likelihood function with no censoring is[4]

$$
\begin{aligned}
L &\propto h(y_{11}, \cdots, y_{ND}) \\
&= \Big(\prod_{i=1}^{N}\prod_{d=1}^{D} f(g^{-1}(y_{id}))\Big)|det(\mathbf{J})| \\
&= \Big(\prod_{i=1}^{N}\prod_{d=1}^{D} f(u_{id})\Big)|det(\mathbf{J})|.
\end{aligned}
\tag{2.19}
$$

The log-likelihood function is

$$
\begin{aligned}
\ln L &= \sum_{i=1}^{N}\sum_{d=1}^{D}[\ln f(g^{-1}(y_{id}))] + \ln|det(\mathbf{J})| \\
&= \sum_{i=1}^{N}\sum_{d=1}^{D}[\ln f(u_{id})] + \ln|det(\mathbf{J})|.
\end{aligned}
\tag{2.20}
$$

2.4 Copula-Based Likelihoods vs. Change-of-Variable Likelihoods

It is useful to compare the copula-based likelihood with the change-of-variables likelihood for the simple case of two duration processes in which there are no covariates, implying $\mathbf{X}\boldsymbol{\beta} = 0$ or equivalently $\theta_d = 1$. To see the relationship between the two approaches, it is sufficient (without loss of generality) to consider the case where there is only one observation point (no spatial interdependence), and therefore I will omit the subscript that indicates observation (unit), and include only the duration subscript. First, consider an SEQ model for logged-duration dependent variables with no covariates, which takes the form

[4] Censoring complicates estimation because the interdependence among durations means the likelihood contains a multidimensional integral. See the next chapter.

$$
\begin{cases}
\ln y_1^* = \alpha_2 \ln y_2^* + \frac{1}{\lambda_1} u_1 \\
\ln y_2^* = \alpha_1 \ln y_1^* + \frac{1}{\lambda_2} u_2
\end{cases}
\Leftrightarrow
\begin{cases}
u_1 = (\ln y_1^* - \alpha_2 \ln y_2^*)\lambda_1 \\
u_2 = (\ln y_2^* - \alpha_1 \ln y_1^*)\lambda_2,
\end{cases}
\tag{2.21}
$$

where $\ln y_i^*$ denotes a dependent variable. y_i^* measures a spell of time and it is assumed to have the FGM Weibull distribution. The Jacobian for $\mathbf{u}$ can be computed as

$$
\mathbf{J} = \begin{pmatrix} \frac{\partial u_1}{\partial y_1^*} & \frac{\partial u_1}{\partial y_2^*} \\ \frac{\partial u_2}{\partial y_1^*} & \frac{\partial u_2}{\partial y_2^*} \end{pmatrix} = \begin{pmatrix} \frac{\lambda_1}{y_1^*} & -\frac{\alpha_2 \lambda_1}{y_2^*} \\ -\frac{\alpha_1 \lambda_2}{y_1^*} & \frac{\lambda_2}{y_2^*} \end{pmatrix}. \tag{2.22}
$$

$$
|det(\mathbf{J})| = \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2|. \tag{2.23}
$$

From equation (2.19), the exact expression of the likelihood derived by the change-of-variables approach is

$$
\begin{aligned}
L = f_{cvt}(y_1, y_2) &= \Big(\prod_{d=1}^{2} f(u_d) \Big) |det(\mathbf{J})| \\
&= e^{u_1 - e^{u_1}} e^{u_2 - e^{u_2}} \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2| \\
&= y_1^{*\lambda_1} e^{-y_1^{*\lambda_1}} y_2^{*\lambda_2} e^{-y_2^{*\lambda_2}} \frac{\lambda_1 \lambda_2}{y_1^* y_2^*} |1 - \alpha_1 \alpha_2| \\
&= \lambda_1 \lambda_2 y_1^{*\lambda_1 - 1} y_2^{*\lambda_2 - 1} e^{-2(y_1^{*\lambda_1} + y_2^{*\lambda_2})} e^{y_1^{*\lambda_1} + y_2^{*\lambda_2}} |1 - \alpha_1 \alpha_2|.
\end{aligned}
\tag{2.24}
$$

The transformation from the second to the third lines of equation (2.24) uses the fact that

$$
e^{u_i} = y_i^{*\lambda_i}. \tag{2.25}
$$

Recall the likelihood function with the joint pdf constructed from the FGM copula, (2.12);

$$\begin{aligned} f_{copula}(y_1, y_2) = \lambda_1 \lambda_2 y_1^{*\lambda_1 - 1} y_2^{*\lambda_2 - 1} e^{-2(y_1^{*\lambda_1} + y_2^{*\lambda_2})} \\ \times [4\alpha - 2\alpha e^{y_1^{*\lambda_1}} - 2\alpha e^{y_2^{*\lambda_2}} + (1+\alpha) e^{y_1^{*\lambda_1} + y_2^{*\lambda_2}}]. \end{aligned} \tag{2.26}$$

By comparing the two likelihoods, (2.24) and (2.26),

$$\begin{aligned} & f_{cvt}(y_1, y_2) = f_{copula}(y_1, y_2) \\ & \Leftrightarrow \alpha = \frac{|1 - \alpha_1 \alpha_2| - 1}{4e^{-y_1^{*\lambda_1} - y_2^{*\lambda_2}} - 2e^{-y_1^{*\lambda_1}} - 2e^{-y_2^{*\lambda_2}} + 1}. \end{aligned} \tag{2.27}$$

In order for the covariance structure that is implied by the copula to be consistent with the structural SEQ relationships among the endogenous variables, equation (2.27) must hold. This equality makes the copula-based estimator for the general SEQ model much more complicated than the change-of-variables-based estimator. Of course, the advantage of working with the copula is that we would have the distribution function. Fortunately, as long as we are working with uncensored duration data, we do not need the distribution function to estimate my models. In the next section I evaluate the change-of-variables FIML estimator against commonly used alternatives.

2.5 Monte Carlo Evaluations of the FIML and Alternative Estimators

Tables 2.1-2.2 present the results of several Monte Carlo experiments in which I evaluate the performance of several estimators: the FIML-SEQ, two-stage least-squares, ML-AEDM, and ML-AIDM estimators.[5] The last two estimators are what most political scientists currently

[5] I also evaluated the three-stage least-squares estimator, but its performance was dominated by two-stage least-squares, so the results are not reported.

use. ML-AEDM stands for maximum likelihood *assumed* exogenous duration model. This is the standard ML applied to a single equation that has an endogenous duration on the right-hand-side. By standard I mean that the estimator treats the endogenous duration as exogenous. ML-AIDM stands for ML *assumed* independent durations model. This is when the analyst fails to recognize that his or her duration of interest is linked in important ways to another or multiple other durations. The ML-AEDM suffers from simultaneity bias while the ML-AIDM suffers from omitted variable bias.

The experimental data are generated using the reduced form SEQ model.[6] I assume two durations, each with an exclusive covariate and a unique shape parameter. More specifically, the structural version of the model is

$$y_1 = \alpha_1 y_2 + \beta_1 x_1 + \lambda_1^{-1} u_1$$
$$y_2 = \alpha_2 y_1 + \beta_2 x_2 + \lambda_2^{-1} u_2$$

These simulations are for the cases of: positive reinforcing interdependence ($\alpha_1 = \alpha_2 > 0$), negative reinforcing interdependence ($\alpha_1 = \alpha_2 < 0$), and mixed interdependence ($\alpha_1 = -\alpha_2$) for small ($N = 100$) and medium-sized samples ($N = 500$). For the naive estimators, reinforcing interdependence should cause inflation bias in the estimated coefficients on the endogenous right-hand-side variables, while mixed interdependence should cause attenuation bias, and the differences in the shape parameters ($\lambda_1^{-1} < \lambda_2^{-1}$) should introduce asymmetries in these biases. The results of my Monte Carlos are mostly as expected. Start with Table 2.1, where I report the results for a small sample (N=100) with positive reinforcing interdependence. The AIDM estimator overestimates β_1 and β_2 as expected. (I get estimates that correspond to the reduced-form parameters instead of the structural parameters.) The estimator also provides inflated estimates of λ_1^{-1} and λ_2^{-1}. The AEDM estimator also inflates

[6] I do not focus on spatial lag models. For those interested, these are studied extensively in Franzese and Hays (2007).

Table 2.1: Monte Carlo Results for Small Sample ($N = 100$), 1000 Trials

Parameter	**Result**	**Independent**	**Exogenous**	**2SLS**	**FIML**
	Mean	-	0.55, 0.93	0.50, 0.49	0.50, 0.49
	S.D.	-	0.01, 0.05	0.06, 0.12	0.02, 0.05
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.05, 0.43	0.06, 0.12	0.02, 0.05
(0.5,0.5)	Mean S.E.	-	0.01, 0.03	0.06, 0.11	0.01, 0.05
	Overconfidence	-	1.19, 1.60	1.04, 1.03	1.04,1.03
	Mean	1.33,1.34	0.97, 0.72	1.00,1.01	1.00,1.01
	S.D.	0.13, 0.17	0.06, 0.12	0.09,0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.36, 0.38	0.07,0.31	0.09,0.17	0.06, 0.12
(1,1)	Mean S.E.	0.12, 0.16	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.07, 1.01	1.05, 1.17	1.01, 1.00	1.03, 1.00
	Mean	2.08, 2.85	0.98, 1.74	- *	1.00, 2.00
	S.D.	0.07, 0.10	0.03, 0.07	-	0.04, 0.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	1.08, 0.85	0.04, 0.27	-	0.04, 0.10
(1,2)	Mean S.E.	0.07, 0.10	0.03, 0.06	-	0.04, 0.10
	Overconfidence	1.04, 1.06	1.01, 1.23	-	1.00, 1.01
	Mean	-	-0.55, -1.06	-0.48, -0.43	-0.50, -0.49
	S.D.	-	0.03, 0.09	0.23, 0.89	0.04, 0.13
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.06, 0.57	0.23, 0.90	0.04, 0.13
(-0.5,-0.5)	Mean S.E.	-	0.03, 0.08	0.23, 0.50	0.03, 0.12
	Overconfidence	-	1.00, 1.14	0.99, 1.80	1.05, 1.09
	Mean	1.34, 1.37	0.96, 0.64	1.02, 1.07	1.01, 1.03
	S.D.	0.46, 0.40	0.13, 0.23	0.29, 0.90	0.14, 0.28
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.57, 0.54	0.14, 0.43	0.29, 0.90	0.14, 0.29
(1,1)	Mean S.E.	0.32, 0.37	0.13, 0.22	0.31, 0.61	0.13, 0.27
	Overconfidence	1.41, 1.08	1.05, 1.06	0.95, 1.48	1.04, 1.03
	Mean	2.48, 2.85	0.97, 1.63	- *	0.99, 1.98
	S.D.	0.34, 0.24	0.08, 0.13	-	0.08, 0.24
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	1.51, 0.88	0.08, 0.40	-	0.08, 0.24
(1,2)	Mean S.E.	0.17, 0.22	0.08, 0.12	-	0.08, 0.23
	Overconfidence	1.95, 1.07	1.01, 1.05	-	1.01, 1.05
	Mean	-	0.41, 0.25	0.53, -0.60	0.51, -0.51
	S.D.	-	0.05, 0.17	0.38, 0.53	0.06, 0.22
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.10, 0.76	0.38, 0.54	0.06, 0.22
(0.5,-0.5)	Mean S.E.	-	0.05, 0.11	0.39, 0.49	0.06, 0.20
	Overconfidence	-	1.00, 1.48	0.99, 1.09	1.05, 1.07
	Mean	0.80, 0.82	0.96, 0.72	1.02, 1.06	1.01, 1.03
	S.D.	0.18, 0.24	0.13, 0.27	0.29, 0.41	0.14, 0.28
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.27, 0.30	0.14, 0.39	0.29, 0.42	0.14, 0.28
(1,1)	Mean S.E.	0.16, 0.22	0.13, 0.23	0.31, 0.41	0.13, 0.27
	Overconfidence	1.10, 1.08	1.05, 1.18	0.95, 1.01	1.04, 1.02
	Mean	1.24, 1.71	0.97, 1.69	- *	0.99, 1.97
	S.D.	0.10, 0.14	0.08, 0.15	-	0.08, 0.24
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	0.26, 0.32	0.08, 0.35	-	0.08, 0.24
(1,2)	Mean S.E.	0.09, 0.13	0.08, 0.13	-	0.08, 0.23
	Overconfidence	1.03, 1.08	1.01, 1.14	-	1.01, 1.03

* The estimates for λ's can be computed using the estimated α and the estimated variance of the error terms. I have not done the computations yet.

Table 2.2: Monte Carlo Results for Large Sample ($N = 500$), 1000 Trials

Parameter	Result	Independent	Exogenous	2SLS	FIML
	Mean	-	0.55, 0.93	0.50, 0.49	0.50, 0.50
	S.D.	-	0.01, 0.05	0.06, 0.12	0.01, 0.05
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.05, 0.43	0.06, 0.12	0.01, 0.05
(0.5,0.5)	Mean S.E.	-	0.01, 0.03	0.06, 0.12	0.01, 0.05
	Overconfidence	-	1.17, 1.59	1.03, 1.01	1.01, 1.02
	Mean	1.33, 1.34	0.97, 0.71	1.00, 1.01	1.00, 1.00
	S.D.	0.13, 0.17	0.06, 0.12	0.08, 0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.36, 0.38	0.07, 0.31	0.08, 0.17	0.06, 0.12
(1,1)	Mean S.E.	0.12, 0.16	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.08, 1.04	1.02, 1.15	0.98, 1.00	1.00, 1.00
	Mean	2.08, 2.85	0.98, 1.74	- *	1.00, 1.99
	S.D.	0.07, 0.10	0.04, 0.07	-	0.04, 1.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	1.08, 0.85	0.04, 0.27	-	0.04, 0.10
(1,2)	Mean S.E.	0.07, 0.10	0.03, 0.06	-	0.04, 0.10
	Overconfidence	1.02, 0.98	1.04, 1.25	-	1.02, 1.01
	Mean	-	-0.56, -1.06	-0.50, -0.49	-0.50, -0.50
	S.D.	-	0.01, 0.04	0.06, 0.12	0.02, 0.05
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.06, 0.56	0.06, 0.12	0.02, 0.05
(-0.5,-0.5)	Mean S.E.	-	0.01, 0.04	0.06, 0.11	0.01, 0.05
	Overconfidence	-	0.98, 1.09	1.04, 1.02	1.04, 1.02
	Mean	1.34, 1.34	0.96, 0.63	1.00, 1.01	1.00, 1.00
	S.D.	0.25, 0.18	0.06, 0.10	0.09, 0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.42, 0.38	0.07, 0.38	0.09, 0.17	0.06, 0.12
(1,1)	Mean S.E.	0.15, 0.17	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.63, 1.07	1.04, 1.05	1.03, 1.01	1.03, 1.00
	Mean	2.60, 2.90	0.98, 1.65	- *	1.00, 1.99
	S.D.	0.20, 0.11	0.03, 0.06	-	0.04, 0.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	1.61, 0.91	0.04, 0.36	-	0.04, 0.10
(1,2)	Mean S.E.	0.08, 0.10	0.03, 0.06	-	0.04, 0.10
	Overconfidence	2.56, 1.13	1.00, 1.07	-	0.99, 1.04
	Mean	-	0.41, 0.21	0.51, -0.52	0.50, -0.51
	S.D.	-	0.02, 0.08	0.10, 0.20	0.03, 0.09
$\hat{\alpha}_1$, $\hat{\alpha}_2$	RMSE	-	0.10, 0.72	0.10, 0.20	0.03, 0.09
(0.5,-0.5)	Mean S.E.	-	0.02, 0.05	0.10, 0.19	0.02, 0.09
	Overconfidence	-	0.98, 1.60	1.04, 1.03	1.04, 1.02
	Mean	0.80, 0.80	0.96, 0.72	1.00, 1.01	1.00, 1.01
	S.D.	0.08, 0.11	0.06, 0.12	0.09, 0.17	0.06, 0.12
$\hat{\beta}_1$, $\hat{\beta}_2$	RMSE	0.21, 0.22	0.07, 0.31	0.09, 0.17	0.06, 0.12
(1,1)	Mean S.E.	0.07, 0.10	0.06, 0.10	0.08, 0.17	0.06, 0.12
	Overconfidence	1.07, 1.07	1.04, 1.17	1.03, 1.00	1.03, 1.00
	Mean	1.25, 1.74	0.98, 1.74	- *	1.00, 2.00
	S.D.	0.04, 0.07	0.03, 0.07	-	0.04, 0.10
$\hat{\lambda}_1^{-1}$, $\hat{\lambda}_2^{-1}$	RMSE	0.25, 0.27	0.04, 0.27	-	0.04, 0.10
(1,2)	Mean S.E.	0.04, 0.06	0.03, 0.06	-	0.04, 0.10
	Overconfidence	1.04, 1.13	1.00, 1.23	-	0.99, 1.01

* The estimates for λ's can be computed using the estimated α and the estimated variance of the error terms. I have not done the computations yet.

the coefficients on the endogenous variables, y_1 and y_2, in this case. These biases, in turn, induce additional biases in the opposite direction for the β_1 and β_2 estimates. Note that the upward bias in the estimate for α_2 is larger than the bias for α_1 because the shape parameter λ_2^{-1} is larger than λ_1^{-1}, and this generates stronger covariance between y_1 and u_2 than exists between y_2 and u_1. The key inequality is $\frac{\partial y_1}{\partial u_2} = \alpha_1 \lambda_2^{-1} > \frac{\partial y_2}{\partial u_1} = \alpha_2 \lambda_1^{-1}$. These patterns are repeated in the experiment with negative reinforcing interdependence. In the case of mixed interdependence, there is attenuation bias in the AIDM estimates for β_1 and β_2. I note that, in the case of the AEDM estimator, the attenuating force is so strong for α_2 that the sign is wrong on average.[7] In all our experiments, the two-stage least-squares estimator performs better than the naive estimators in terms of bias, but occasionally, particularly in the small samples, performs worse in mean-squared-error terms. The two-stage least-squares results are much better for the medium-sized sample.

The FIML estimates are virtually unbiased in all cases even in the smaller sample, and the standard error estimates, calculated with the observed information matrix, are accurate. The standard error accuracy of the other estimators is frequently poor. The AIDM estimator is always overconfident, and the degree of overconfidence increases with the true variance of the sampling distribution. With the AEDM estimator, the degree of overconfidence is higher for the more badly biased coefficients (i.e., α_2 and β_2), a particularly disturbing combination that makes sound inference difficult.

[7]Technically, the bias need not be attenuating in the sense that the estimate is, on average, closer to zero than the truth. Small effects in one direction can be overwhelmed by bias inducing forces pushing in the other direction so that the estimate, on average, has the opposite sign and is farther from zero than the truth.

2.6 Application: Democratic Transitions and Survival in Africa

I illustrate these methodological findings in a study of the determinants of democratic transitions and consolidation in Africa. What makes some formerly non-democratic countries more likely to undergo democratic transitions than others? What makes some countries remain democratic while others revert back to dictatorships? There have been longstanding scholarly interests in the empirical studies of regime transitions and the stability of democracies. Among both theoretical and empirical analyses of democratization and consolidation, the most conventional approach is to model the latent probability of a country's regime change from one type to another as a function of structural predictors, given a country and a year (e.g. Boix 2003; Przeworski et al. 2000; Epstein et al. 2006). By structural predictors/variables, I mean country-specific political, economic, and social attributes, such as the level of economic development, inequality and ethnic fragmentation. A newer approach to this question is to model directly the spell of time till a transition occurs, either to or from democracy (e.g. Alemán and Yang 2011). For example, for models that examine the timing of democratic reforms, the dependent variable in these studies is measured as a spell of time between the beginning of transition and the emergence of a democracy. Likewise the survival of democracies is measured as a spell of time between the start and end of democratic regimes.

Despite the extensive empirical literature that attempts to explain the timing of democratization by various structural variables, important predictors seem to be missing. These are mainly behavioral and actor-oriented contributors to regime transitions. In the following, I highlight three kinds of these behavioral and strategic factors that are at least as important as country-specific attributes in determining the timing of regime transitions. They are (i) the elite's expectation about the durability of the future democracy, (ii) the democratic opposition's anticipation about the effectiveness of its anti-dictator movements, and (iii) the potential effects of democracy-appreciating civic cultures that can grow as liberalization slowly progresses.

2.6.1 Interdependence between liberalization and the Survival of Democracies

Some recent theoretical works touch upon the issue (i). For example, actor-oriented theories (namely game theoretic models) of democratization (e.g., Acemoglu and Robinson 2006*a*) claim that the authoritarian elite could form an anticipation about the durability of the future democracy, or similarly the cost of staging a coup in the future to reverse the democracy. This anticipation by the elite about the survival of the potential democracy affects their decisions on when to democratize the country. If they expect that it would be costless to reverse the democracy by a coup, the elite might not negotiate hard with the democratic opposition, making the transition process quicker. On the contrary, if they expect democracy to last longer once it emerges, then they might negotiate harder with the opposition, prolonging the transition duration as a consequence. Østerud (2011) also provides a theoretical argument about the elite's anticipation, based on cases of post-communist regimes in east and central Europe. He claims that authoritarian leaders are more willing to negotiate with the democratic opposition, when "they conceive this as a lesser or less risky evil" (p.48); in other words, authoritarian leaders are more willing to negotiate with the democratic opposition, when they sense that they might be able to retain some political power and influence after the transition. A specific example would be "the potential cost of coup" that the elite could stage to regain political power. This strategic thinking by the authoritarian elites can induce a positive causal effect from the expected duration (durability) of a future democracy to the length of the current dictatorship. The longer (expected) survival "causes" a prolonged dictatorship period.

It is not only the elite that is concerned about the future regime outcome. The democratic opposition can also forms anticipation about the success and effectiveness of its anti-government movements. For example, citizens living in dictatorship attempt to gauge how likely a collective action against the authoritarian government would succeed. They might simply sense the odds of success, or they might learn from an experience of neighboring countries (Weyland 2009). The movements can also occur within the government. The opposition force within the government might gauge how much of political power (relative to the authoritarian incumbent) they possess to push for a new democratic constitution. For example, Seely (2005), comparing Benin's successful case and Togo's unsuccessful case of democratization, argues that the pro-democracy group's (including both governmental and civic actors) relative political power facilitated Benin's democratization. However, in my view, it is also important to note that the democratic opposition does not usually confront against the incumbent authoritarian regime, just because they know they have greater political power relative to the incumbent, because if the liberalization movement or negotiation does not lead to a lasting democracy in the future, they would most likely be punished by the continuing authoritarian government. This is a vital component of the classical collective action dilemma with the presence of severe punishment. The same applies to the theory of economic crises and democratization. Gisselquist (2008), for instance, partially attributes Benin's successful liberalization to the increasing pro-democracy movements in the late 1980's. She argues that those movements and the National Conference that eventually led to a new democratic constitution were triggered by economic crises. To me this is only half convincing. It is because, again, the opposition would not participate in costly collective actions unless they anticipate a higher probability of success in the future. In fact, Nigeria in the late 1980's was also a dictatorship that was experiencing a major recession, but liberalization in Nigeria did not occur for another decade. In each of these cases, unobservable strategic calculations and assessment by the opposition involved their expectation about how likely their costly political actions would lead to the emergence and ideally the stability of a democracy.

Take another case in the Eastern and Central European countries in the late 80's and early 90's. After five years of Mikhail Gorbachev's leadership in the Soviet Union, the direct rule of the Soviet communist party came to an end and the effect of Soviet power started to decline in east European countries. With the stagnating economy and the weakened political authority of the centralized communist regimes, the liberalization demand from these societies heightened. There were increasing numbers of major and minor uprisings by the opposition. Lewis (2000) points out that gain in the political power by the liberalization forces was the key in the course of this change in the political atmosphere. He claims that the relaxation of dictatorship and repressive practices after Stalin's death in 1953 was not enough of liberalization of the regime, but it was "sufficient to permit freer communication and a degree of contact with the west that only made awareness of relative failure of the communist system that much sharper" (p.13). This awareness of the economic and institutional failure of the communist regime in general invigorated the opposition force toward the end of the 80's. We, researchers, cannot observe exactly what triggers the opposition's "awareness" for the promising political atmosphere, but the key political dynamic one should note is that, to some extent, the opposition can sense the expected effectiveness of their movements and the potential emergence of a democracy in the future, and it affects the actual occurrence of revolutionary uprisings, shortening the duration of the dictatorial regime. To summarize the essence of these cases in Africa and Europe, the longer (expected) survival of democracies can "cause" a shorter dictatorship period through heightened political activities by the democratic opposition. Note that the causal direction is the same as the first case, but the effect is opposite.

Finally, there is also a dynamic that manifests the opposite causal relationship. A prolonged period of an authoritarian regime could also mean greater opportunity for the society to nurture democratic values before democratization takes place. Obviously, the possibility of such a phenomenon depends on how active and organized the democratic opposition is. As the behavioral literature suggests, this socio-political culture that potentially prepares the citizens to value democracy stabilizes democratic systems later once such a regime is established (Almond and Verba 1965; Putnam 2002). There is also a different set of theories that suggests the same direction of causal effect, coming from the "gradualism" debate.

This argument suggests a positive causal effect of the liberalization duration on the survival of democracy.

What these theoretical claims–either formally or verbally–suggest is that, in predicting the timing of democratic transitions (including whether or when they occur), the potentially co-existing causal relationships between liberalization and consolidation are non-ignorable contributors. Note that the first two theories suggest the same causal direction (from survival to democratization) but they propose competing effects–positive and negative causal relationships. The third theory suggests the opposite causal direction.

2.6.2 Definition: The Liberalization Duration and the Survival of Democracies

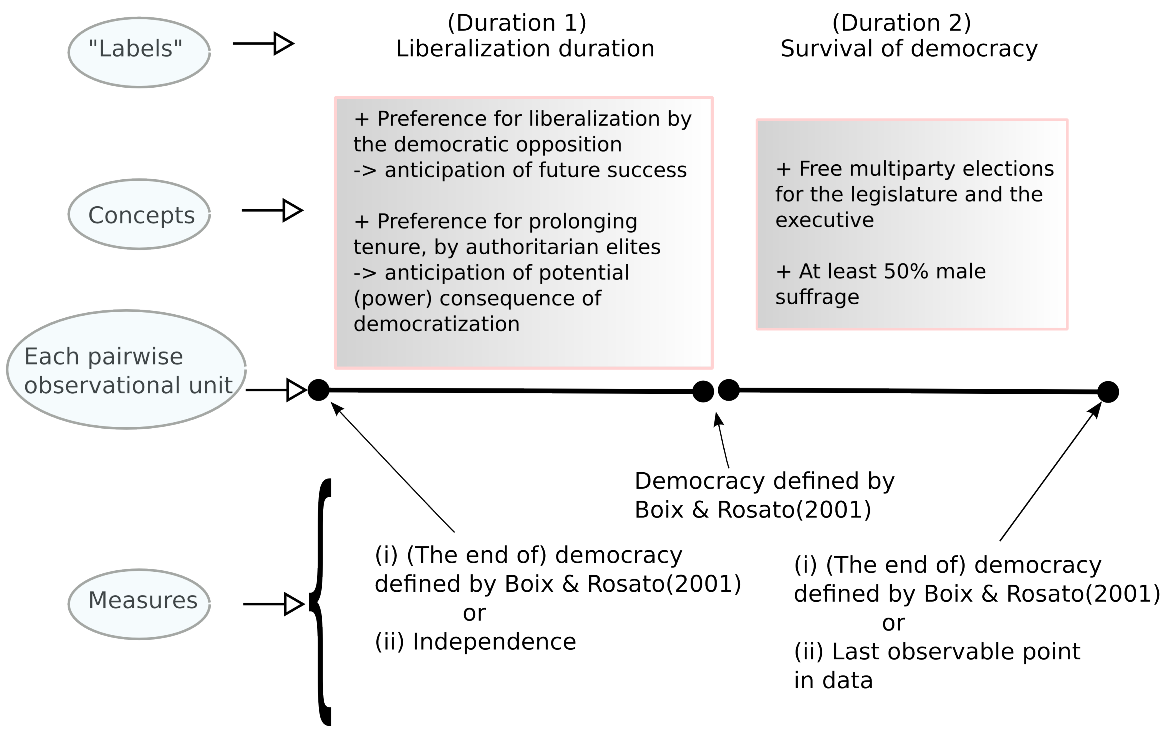
The previous section discussed three main mechanisms through which one should expect dependence between the liberalization duration and the survival of democracies. To test for the existence of such dependence, first I need to define what constitutes each of the two processes and discuss how to measure the two durations such that they reflect the aforementioned theories.

First, for the survival of democracies, I subscribe to the concept built in Boix and Rosato (2001), and also used in Svolik (2008). Boix (2003) summarizes these criteria for democracy as follows. A country is a democracy if: (i) the legislature is elected in free multiparty elections; (ii) the executive is directly or indirectly elected in popular elections and is responsible either directly to voters or to a legislature elected according to the first condition; and, (iii) at least 50% of adult men has the right to vote. A spell of a democratic regime should start when all of these institutional conditions are met, and the period potentially ends when at least one of these conditions is violated.

What is more challenging is the definition of liberalization: there are many different ways to define liberalization in different reserach tradition. In short, in my study, the "liberalization" period is equivalent to the period of non-democracy. The reason is as follows. The theories of interdependence imply that, the end of liberalization can occur when either the authoritarian elite or the democratic opposition takes some action that leads to a change in the status quo—namely the emergence of a democracy based on the above definition. Before any of these political actions take place, the elite is concerned about the liberalization forces from the opposition, assessing whether it is possible to prolong its regime without being overthrown by an uprising. The elite is forced to make this assessment to some extent throughout its tenure. This is a baseline assumption of formal theoretic models of regime transitions (Acemoglu and Robinson 2006*b*; Wintrobe 1998). At the same time, during the liberalization period, the democratic opposition is concerned about the likelihood and opportunities to change the status quo to achieve a democracy. Therefore, the liberalization phase, at least in this study, should be defined as a period in which (i) the authoritarian elite is concerned about prolonging its tenure, and (ii) the democratic opposition (in- or outside of the incumbent government) is concerned about the effectiveness and momentum for anti-authoritarian movements that can lead to a democracy.

To measure the duration of such periods, we need to know when these actors start to be concerned about these future political outcomes. In fact, both of them do so as soon as a non-democratic regime starts. For authoritarian elites, the balance between repression and avoidance of the major headwind of liberalization forces is a day-to-day political matter. For the democratic opposition as well, they are the "opposition" as soon as a non-democratic regime emerges and their calculations for exiting the status quo then begins. This is the reason I define the period of liberalization as a spell of time when the regime is non-democratic. This can be measured as a period between the end of the last democracy and the beginning of the next democracy, based on Boix and Rosato (2001)'s definition, or from birth of non-democracy to establishment of democracy. Figure 2.2 schematically summarizes the definition of both durations used in this paper.

Figure 2.2: Concepts and Measures of the Two Dependent Variables



One might think that the liberalization phase is when the both parties are in an explicit negotiation stage. For the purpose of my study, it is not, because for an explicit negotiation to take place, there needs to be an explicit decision by the opposition (or the international community in some cases) to propose such a conference to discuss potential democratic reforms with the authoritarian incumbent. In the framework of my study, this is already an intermediate political *action* or *outcome* that materialized after the two parties' strategic calculations, which involve their anticipation about the future regime outcomes. One could potentially study regime transitions as a two-stage event, where decisions on proposing/demanding a national conference takes place first, and then the final political regime outcome is determined.[8] However, this is beyond the scope of my research question and empirical analyses in this paper.

2.6.3 Africa as a Sample

My empirical study focuses on countries in Africa. The first liberalization event starts in 1956 with the independence of Sudan. The data range from 1956 through 2001. My choice of the sample is largely driven by the data availability, but in a very particular way that stems from dealing with "duration" variables.

Since many democracies have survived for a long time and some of them are still surviving at this moment in 2012, the second duration—the survival of democracies—is always cut off (*censored*) to some degree in any data we have. This truncation is inevitable.[9]

However, even the "left side" of the data, namely the beginning of liberalization periods, can also get truncated in many cases. There are two reasons for this. First, in some cases, it is difficult to measure the start of the liberalization stage in a consistent manner across countries. Before democracy (in Boix and Rosato (2001)'s sense), there were long periods of non-democracy in the US and UK, for example.[10] The beginning of such periods were the emergence of those states itself. However, it is difficult to compare those "emergence

[8] In fact, Gisselquist (2008) states that Benin's transition in 1991 had three stages: first, the collapse of the authoritarian government due to economic crises; second, the negotiation at a National Conference first about the economic situation and eventually about a new democratic constitution; and third, the consolidation stage. Bratton and van de Walle (1992) also present a 2x2 table with countries in Africa that did and did not experience protests and countries that did and did not achieve democratic reforms.

[9] I discuss a potential statistical treatment for right-censored observations in Essay2. The treatment improves the statistical results, but it never "fills in" missing observations due to censoring.

[10] Alternatively, but without losing the validity of the point made here, one could argue that the UK experienced a long period of non-democracy and then a long-period of partial democracy before a full democracy emerged.

periods" with The independence of African states. Second, even if one can define the start of liberalization spells for all countries, the majority of data for other covariates is available typically only for the post-WWII era. Alemán and Yang (2011), in their cross-national study with 77 countries, states that most of their covariates are available only from 1970, limiting their empirical study for the period between 1970 and 1999. For these reasons, and any liberalization experiences that start before 1950 would most likely be cut off in my analysis, obscuring the true durations of liberalization. Even worse, if these countries democratize and they are still democracies at the end of the data range (year 2001), then the duration observations would be cut off left and right. This makes my duration analysis very unreliable.

By focusing on countries in Africa, I can alleviate the potential problem occurring from the "left side" of the data set. This is because, first, the natures of independence in these countries are similar, and hence the the natures of liberalization durations are comparable. But more importantly, these countries became independent mostly in the late 1950's and the 60's, which allows me to include the whole liberalization periods, without being interrupted by the data availability for other covariates.

There are a number of other empirical studies focusing on a certain geographical area. Many of them attempt to justify their limited sample by claiming that it would control for unwanted and unobservable differences across regions. For example, I could claim that I can eliminate the cross-region differences in the level of ethnic fragmentation (higher in Africa) and the level of institutional development (lower in Africa), by studying only Africa. However, this justification is not convincing to me. It can also be done by including regional dummies in regressions, maintaining a larger sample size, which can be more preferable from the viewpoint of statistical inferences. The problem associated with the data availability and duration variables, which I explained above, however, does not go away by including region dummies. Therefore, I decided to start my empirical analyses from this particular set of countries.

Obviously, my approach is vulnerable to the criticism about the external validity of the empirical results. Do my findings apply to other regions? Nevertheless, I believe this is a reasonable first step to make a contribution by conducting empirical tests of the *suspected, but not yet tested*, interdependence between two important political processes.

2.6.4 What the Interdependent Duration Approach Can Offer

We have not yet seen systematic empirical studies that convincingly takes into account the interdependence between the emergence and the survival of democracies. For example, even though Acemoglu and Robinson (2006*a*) demonstrate an intriguing model that connects elites' strategic decision-making in the democratic transition and consolidation phases, the only empirical statement made is "this result may help to explain the existence of highly redistributive but relatively short-lived populist regimes in Latin America." Alemán and Yang (2011) employ a duration model to explain the timing of transitions and consolidation, including both the typical structural variables and other variables that could capture the costs that both the elite and the masses might face in their strategic movements. However, they model the liberalization duration and the survival of democracies in isolation as two separate models. Due to this construction, the models fail to estimate the existence and the magnitude of potential causal relationships between the two dependent variables. As can be seen in my empirical results later, the independent approach leads us to qualitatively different conclusion about the effects of durations.

My new interdependent duration model can test for the existence and the magnitude of the causal relationships, as well as the effects of traditional predictors for the democratic transition and consolidation. I explore the use of a multivariate duration model described in this essay to unify the two political processes, and build an empirical model that incorporates suggested, and yet untested, theoretical foundations in the studies of democratization. I believe this is a very useful attempt, in that it can reveal the existence and the strength of actors' strategic thinking (which is usually unobservable for researchers) from observable data.

2.6.5 Variables and Data

Dependent Variables

My dependent variables are two durations: the first is a transition duration from being either ruled by a Western European country as a colony or being an independent autocracy to being democracy, and, the other is the survival of a democracy. In this application, I refer to the first type of duration mainly as the *liberalization* duration, and the second as the *survival (of democracy)* duration.

The liberalization duration is a spell of time in which countries are not democratic. There are two situations that belong to this type of periods among the states in Africa. First, if a country was not independent, the liberalization duration is conceptualized as years between its independence and democratization. The second case starts when a country's democratic regime ends. In turn, the liberalization duration is conceptualized as years between the end of democracy and the beginning of the next democracy period.

The survival duration is a duration of "democracy": years between the beginning and the end of a democracy. If a country is democracy (not reverted to autocracy) in the last observed time period in the data set, then the case is said to be *right-censored*, and a democracy spell is simply measured as years between the beginning of the democracy and the last observed period in the data.[11]

For the survival of democracies (the second duration), I use data compiled by Svolik (2008), based on the original Boix and Rosato (2001) data but extended to the year 2001. Among

[11] The methodological issue related to right-censored cases is discussed in Essay 2. Essay 2 suggests a way to model the censoring information in the interdependent duration modeling approach.

my sample of African countries, there are 32 spells of democracies, and 17 cases are right-censored; i.e., there are on-going democracy spells in the last observed period (the year 2001) of the data set in 17 cases.[12]

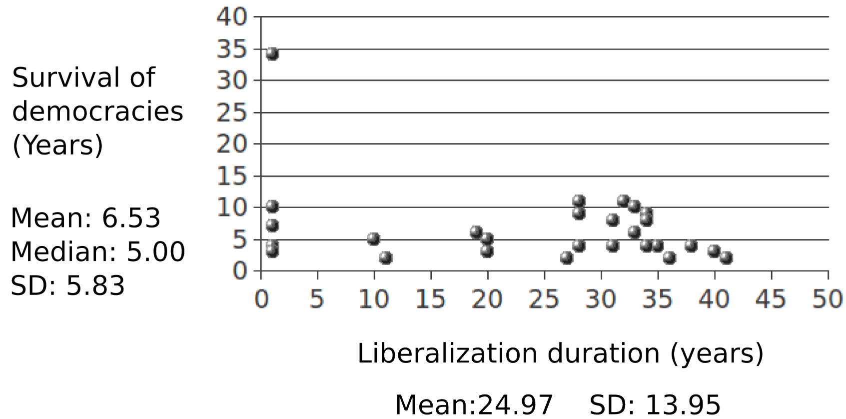
For each of the democracy spells, I constructed the measure for the liberalization spell, which comes before the democracy spell. If the democracy spell is the first one for the country since its independence, the liberalization duration is the number of years between the independence and the start of the democracy spell. If it is not the first democracy spell for the country, then the liberalization duration measures the number of years between the end of the last democracy spell and the beginning of the next one.

Even though half of the survival spells are right-censored in my data, I was able to avoid artificial left truncation that stems from data availability, by focusing on countries in Africa. This is due to the fact that many of them became independent roughly around the same time in the 1950's and 60's, and also because relatively reliable measures of other covariates are available for the post-war period. For example, the advantage of Alemán and Yang (2011)'s approach is that they incorporate many more countries (77 countries) maintaining a fairly large sample size. However, due to the data availability for other covariates, they had to focus on the data ranging between 1970 and 1999. This implies that not only the democracy spell but also the liberalization duration for most countries in their model are cut off severely.

As summarized in Figure 2.3, the liberalization duration ranges from 1 to 41 years, with the average of about 25 years. The typical variation in the liberalization duration is about 14 years, showing a wide range of transition durations. There are 6 cases where transitions occurred within a year from their independence, meaning that democratic institutions were established as soon as the countries became independent. Out of these 6 countries, 5 of them

[12]There were 4 more democracy spells included in the original data; however I had to disregard them due to non-availability of other covariates for these cases.

Figure 2.3: Scatterplot of the Liberalization and Survival Durations



have experienced authoritarian reversals. The survival duration, including the right-censored cases, ranges from 2 to 34 years with the average of 6.5 years and the standard deviation of 5.8 years.

In the figure, one might notice a data point that is located far out of the way of the main clustering. This is Mauritius, with the liberalization duration of 1 year and the survival of 34 years. The 34 years of survival is much longer than other observations. This is why the figure also reports the median survival (5 years). Mauritius is one of the most economically successful countries in Africa, and its economic development level seems to be strongly associated with the longer survival of its democratic political system. In the main estimation results reported later in this essay, I keep Mauritius in the data, but I also discuss how keeping this case in the data does *not* change the qualitative results in terms of the duration dependence and the estimates of most of the other covariates. Only the economic variables become weaker or statistically insignificant in some cases without Mauritius.

Lastly, one should note two things about these duration variables. First, as explained in the methodology section, the dependent variables used in the estimation are the logged form of these raw durations measured in years. Second, in my theory, these durations are not only the dependent variables, but also the explanatory variables of each other. As an explanatory variable, a duration variable is measured in years, and not as the logged form.

Other Covariates

It is not possible to make any meaningful statements about the bivariate relationship between the two durations without controlling for various country– or case–specific variables, and also for potential simultaneous influences between the durations. In this section, I describe other covariates included in the model and how they are measured.

My selection of control variables follow the literature, mainly Przeworski et al. (2000). It allows us to make my theory somewhat consistent and comparable to what the existing research on democratization and democratic consolidation presents. I test for the effects of the following variables. I first describe time-variant variables, whose values potentially change across time periods. For these time-variant variables, I take data from the year of transitions. For liberalization, I use values from the year in which a transition to democracy occurs. This is the end of the liberalization period. For survival, I use values from the year in which a democracy collapses. In cases where observations are right-censored at year 2001, I take the measures for 2001.

Real GDP per capita. As Lipset (1959)'s social and economic requisites theory and many other empirical studies of democratization predict, a country's economic circumstances can be highly correlated with the occurrence and timing of democratization (Huntington 1991; Boix 2003; Boix and Stokes 2003). Przeworski et al. (2000) maintain, on the other hand, that a country could achieve democratization in any economic environment, predicting no effect of the GDP level on the occurrence of democratic transitions. They also theorize and demonstrate empirically that the economic development is important to sustain a democratic regime once it is established. The empirical findings on the economic variables have been mixed in the literature, but it is nonetheless crucial to include this variable in my model of democratic transitions and consolidation. To allow the possible curve-linear relationship between the economic development and the democracy levels, I start with a model specification that includes both *real GDP per capita* and the squared term, *(real GDP per capita)*^{2}. Theoretically, it makes sense to claim that the marginal effect of economic development on democratization diminishes as the development level becomes very high (hence the effect is curve-linear); however, it is not implausible that the odds of democracies' survival only increase as the GDP level increases (hence no need for the squared term). I will test for both possibilities in my empirical analysis and compare the model fit indices. I use real GDP data from Gleditsch (2002). It measures the real GDP per capita in hundred-thousands of constant U.S. dollars with the base year being 1996.

Fuel export rate. Another economic variable is *fuel export* that measures the percentage (in ten percent) of a country's fuel export of all the merchandise exports. The data are taken from the World Development Indicator (*World Development Indicators* N.d.). Democratization implies the government's enfranchising the poor or the repressed. This moves the median voter position from somewhere in the wealthy group to a point in the then repressed group–the class of citizens that more likely chooses to impose higher tax rates for the elite than in the former authoritarian regime (Boix 2003). While owners of relatively mobile businesses, such as manufacturers, could move their production sites outside the country if higher taxes are implemented, natural resources are highly location-specific assets and, making it more difficult for the rich to "move" such assets when the tax rates increase (Boix 2003). Therefore in the fuel-rich countries, the elites have an incentive to block democratization movements, making the liberalization process slower.

Urban population rate. I expect that the urban population rate would be positively correlated with democratization for a couple of reasons. First, the urban population is more likely exposed to foreign culture, among which can be democratic countries. Second, higher degrees of urbanization imply higher population concentration. This could facilitate, in conjunction with the exposure to more democratic foreign culture, solving the collective actions that are oftentimes a key to overturn the dictatorial incumbents either by voting for the democratic opposition or revolting against the government. The data comes from the World Bank's World Development Indicators (*World Development Indicators* N.d.). It measures "the percentage of a country's urban population living in that country's largest metropolitan area." In my estimation, the unit is in ten percent.

The remaining covariates are time invariant. These are characteristics associated with each country that are constant across time periods.

Military regime. This is a dummy indicating whether the authoritarian elites in the liberalization period were military personnel or not. The data comes from Svolik (2008) and the study finds that having a military dictatorship prior to a democratic regime significantly decreases the odds of democratic consolidation, but does not have a significant influence on the odds of authoritarian reversals for transitional democracies.[13] Regarding the association between military dictatorships and the prospect of democratization, Geddes (1999) argues that military juntas can be least resistant to liberalization due to its concern about the internal integrity as a political actor. However, there is also a possibility that military dictatorships have more direct control over the civic society than any other forms of dictatorships and hence are more resistant to liberalization.

Not independent. The survival-of-democracy process includes this dummy, which indicates cases where democracy spells start immediately after the country's independence. In my liberalization duration variable, these cases are coded to have the transition period of 1 year instead of 0 year, but these cases had essentially no experience of dictatorships directly before their democracy periods. The dummy is necessary because theoretically there is no possibility for these cases to have any potential effects of characteristics of the liberalization period on their democratic consolidation. There is no way for the civic society to learn to nurture democratic values, or for authoritarian elites to contemplate on the possibility of

[13] Svolik (2008) also includes the civilian and monarchy dummies, so the category of dictatorship is more fine grained in the said study.

regime reversals in the future. Therefore, I need to include this dummy to capture any effect that might make these cases distinct from others.

Commonwealth. This variable is a dummy indicating the commonwealth membership. This variable is to capture the possible legacy of British colonization. Due to the imposed British political institutions during the colonial era, these countries might have some political attributes in common with Britain that might lead to stable democratic regimes, regardless of their own political history before or after the colonization.

Muslim population rate. Following Przeworski et al. (2000), a religious profile of countries is considered. It measures the time-invariant percentage of the population (in ten percent) that belongs to the religious group. The data are mainly from Przeworski et al. (2000) and for the countries that are not included in Przeworski et al. (2000), I took data from the current issue of the CIA World Factbook (Agency Date accessed: July 15, 2012). In this study, this religion dummy serves mainly as a control variable for a social aspect of the countries.

Presidentialism. The type of executive is one of the most debated political institutions in terms of its effects on the probability of democratic consolidation. For instance, Linz (1994) points out the fragility of democracies under presidentialism, while Cheibub (2007) argues that the seeming negative effect of presidentialism is a mere artifact of the possible effect of term limits that presidents often face, and there is no direct effect of presidentialism on the survival of democracies. I include the presidentialism dummy to investigate its potential effect on the survival.

Another covariate that is missing from my current analysis but should be considered, particularly for countries in Africa, is the amount of *per-capita foreign aid.* For instance, among the studies on democratization and Africa, Gabizo (2005) argues what made Benin's democracy more stable than that of Niger's, even though the two countries are similar in many other political economic aspects, is the higher amount of per-capita foreign aid Benin received in the 1980's. Gisselquist (2008) also confirms this by comparing Benin's aid amount to that of Niger, Nigeria and Togo. Empirical findings seem to be divided in the foreign aid literature. Brown (2005), summarizing an extensive literature of the democracy promotion and consolidation, claims that foreign aids is more effective for transitions but less effective for consolidation. A recent finding by Dietrich and Wright (2011), however, is that democracy aids—when they are separated from general economic aids—have a positive influence on different aspects of democratic consolidation.

2.6.6 Estimation and Results

Table 2.3 presents three sets of results obtained by the interdependent duration model (Model (1), (2) and (3)), and a set of results obtained from a conventional univariate Weibull duration model (Model (4)). The coefficients in Model (4) were estimated separately for the effects on the liberalization duration and on the survival of democracies.[14] Model (1), (2) and (3) are some of the most statistically preferred specifications based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and Model (3) is the most preferred.[15] In the following, I briefly go over the estimation results focusing on Model (3)–the most statistically preferred specification, estimated by the interdependent duration model, which allows for the possibility that both durations can be the causes and consequences of the other durations.

Main Effects: Liberalization and Consolidation are Dependent on Each Other

The results are quite striking. Regardless of the model specification (whether using Model (1), (2) or (3)), both the liberalization speed and the duration of democratic regimes are highly interdependent on each other, in both ways. The coefficients of "Survival duration

[14] In other words, the top half of Model (4) in the table is from a model that explains the duration of democratic transitions, and the bottom half is from a model that predicts the survival of democracies.

[15] AIC and BIC scores for each specification is reported at the bottom of the table. Specifications with lower AIC or BIC scores are more preferred.

(α_1)" and "Liberalization duration (α_2)" reveal these effects and it is noteworthy that the statistical significance is high for all the models even with the small sample size.

First, focusing on the top half of the table, the anticipated survival of democracies has a *negative* effect on the duration of the liberalization period. This means that, when a durable democracy is expected in the future, it tends to take shorter for the country to democratize, while when a fragile democracy is anticipated in the future, it tends to take longer for the country to democratize. Why would these relationships hold?

As I discussed earlier, there are at least two important ways the anticipated durability of democracies can influence the timing of democratic transitions. One is the elite's anticipation. When they anticipate a fairly low cost of staging a coup in the future, the elites might be more willing to democratize. However, this theory predicts a positive causal effect of the survival on the transition duration. My empirical results suggest that this is not a prominent causal relationship manifested in the story of democratization. What is more plausible, from the sign of the estimate, is the anticipation by the democratic opposition, either at the civic or the elite (party) levels. When they highly anticipate that overturning the authoritarian government will be successful and effective, the momentum against the incumbent government can grow fast among citizens or political parties. This can lead to a successful movement that leads to an influential uprising or a democratic constitution, resulting in a shorter duration of the authoritarian regime.

This is not necessarily to claim that the elite's anticipation does not exist. There can be both the effects of the elite and the opposition's anticipation, but the estimate tells us that the effect of anticipation by the opposition is much more prominent than the other. One way to interpret the estimate is to see it as a "net" strength of the anticipation effect by the democratic opposition, compared to that of the elite.

The magnitude of the effect is also substantial. A one-year increase in the anticipation of the survival duration causes the liberalization duration to decrease by about 4 years.[16] Obviously, it is not plausible to argue that the democratic opposition fine-tunes the effectiveness of its movements with the precision implied by the metric of years of survival, but this results highlights the strength of the effect working in the background.

Second, focusing on the bottom half of the table, the length of the liberalization period has a *positive* effect on the survival of democracy, implying that, after a longer democratization period, a democracy tends to last longer. This means that, when a democratic transition is slow, it tends to lead to a longer-living democracy. There can be various mechanisms working behind this causal relationship. From the behaviorists' point of view, this can imply that cultures that nurture democratic values do grow during the process of liberalization, stabilizing the future democracy if the transition occurs. This can also be consistent with the prerequisite theories of democratization. A longer liberalization period could provide the country with more time to develop institutional foundations that eventually stabilizes a democracy, if a transition occurs in the future.

The magnitude of this effect is smaller than the other causal relationship, but it is still substantial with a great statistical significance. Each one-year increase in the liberalization duration causes the democracy that comes later to survive 2.5 years longer.

The Effects of Covariates

In the most preferred model (4), the economic variables, real GDP levels and its squared term (only in the liberalization equation), turn out to affect both durations. As expected, the real GDP level has a negative effect on the duration of liberalization; in other words, the liberalization process becomes faster as the economic development level increases. Also the marginal effect of the development level on liberalization diminishes as the level goes higher, suggesting the curve-linear relationship well-discussed in the literature. For the survival of

[16] Note that the dependent variable is logged.

democracy also, the economic indicator has a positive effect, suggesting a strong association between the development level and the stability of democracies. However, I should note that the effects of economic variables seem to be largely driven by a single case–Mauritius, which became a democracy immediately after its independence from the United Kingdom and was consistently democratic until the last observed period in the data. I discuss this issue in the next section.

An important political institution variable is military regime. Contrary to Geddes (1999), I find that military dictatorships tend to last longer than other forms of dictatorships. My finding is also consistent with Alemán and Yang (2011), and I suspect that this is due to the power of the direct (physical) control unique to military regimes. I did not find a statistically significant result for the effects of a military regime on the survival of the following democracy. However the sign of this coefficient (in any of the three specifications) is consistent with Svolik (2008), suggesting that having a military dictatorship prior to a democracy makes the democratic regime more fragile.

The second important institutional variable is the effect of presidentialism on the consolidation of democracy. The result of my main model (4) is consistent with the finding in Cheibub (2007), showing no effect of presidentialism on the stability of democracies. However, it is noteworthy that when this coefficient becomes (slightly) significant depending on the model specification (see Model (1)), the sign of the coefficient is in fact consistent with Linz (1994), indicating that presidential democracies are more fragile compared to other type of executive institutions.

Possible "Outlier": Mauritius

As one can see in Figure 2.3, Mauritius is a seeming outlier in terms of its long survival duration. I also estimated the models excluding Mauritius. The qualitative results of all

Table 2.3: Estimation with Duration Models of Liberalizing and Autocratizing Transitions

	Simultaneous Duration Models[a]			Univariate Models[b]
	(1)	(2)	**(3)**	**(4)**
Dependent variable: Liberalization duration ($y_1 = \ln(y_1^*)$)				
Real GDP/cap	-0.116**	-0.080	-0.100*	-0.073
	(0.056)	(0.067)	(0.060)	(0.050)
(Real GDP/cap)^2	0.001**	0.001	0.001*	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Commonwealth	-0.051	0.083	0.023	-0.051
	(0.395)	(0.522)	(0.457)	(0.327)
Urban pop rate	0.245	0.270	0.261	0.213
	(0.166)	(0.227)	(0.196)	(0.138)
Muslim pop rate	-0.092	-0.122*	-0.110*	-0.038
	(0.050)	(0.066)	(0.057)	(0.041)
Military regime	0.648*	0.818*	0.754*	0.470*
	(0.339)	(0.473)	(0.403)	(0.282)
Fuel export rate	-0.467	-0.109	-0.080	-0.029
	(0.064)	(0.082)	(0.071)	(0.057)
Intercept	5.379***	6.279***	5.893***	3.683***
	(0.771)	(0.938)	(0.816)	(0.613)
Survival duration (α_1)	-1.060***	-1.789***	-1.458***	-0.076***
	(0.167)	(0.191)	(0.004)	(0.028)
Shape parameter 1 (λ_1^{-1})	1.266***	0.958***	1.092***	0.424***
	(0.203)	(0.156)	(0.161)	(0.158)
Dependent variable: Survival duration of democracies ($y_2 = \ln(y_2^*)$)				
Real GDP/cap	-0.041	-0.010	0.012***	0.011**
	(0.033)	(0.027)	(0.004)	(0.004)
(Real GDP/cap)^2	0.0003*	0.0001		
	(0.0002)	(0.0002)		
Commonwealth	0.021	-0.071	-0.128	-0.159
	(0.245)	(0.229)	(0.227)	(0.240)
Urban pop rate	0.089	-0.044	-0.068	-0.009
	(0.130)	(0.116)	(0.104)	(0.106)
Muslim pop rate	0.007	-0.020	-0.025	-0.033
	(0.040)	(0.034)	(0.033)	(0.032)
Military regime before	-0.940***	-0.359	-0.272	-0.068
	(0.307)	(0.244)	(0.241)	(0.267)
Not independent before		2.986***	2.593***	0.249
		(0.333)	(0.328)	(0.603)
Presidentialism	-0.470*	-0.089	0.014	0.053
	(0.277)	(0.256)	(0.237)	(0.246)
Intercept	1.667***	-0.712*	-0.411	1.879***
	(0.497)	(0.420)	(0.370)	(0.590)
Liberalization duration (α_2)	0.403***	0.948***	0.785***	0.001
	(0.063)	(0.0002)	(0.00003)	(0.020)
Shape parameter 2 (λ_2^{-1})	1.700***	2.054***	2.083***	0.731***
	(0.246)	(0.298)	(0.302)	(0.140)
Log-likelihood	-66.11	-48.78	-51.89	-40.37/-26.59
AIC / BIC	172.23 / 201.54	151.32 / 182.11	143.78 / 173.10	

Note: N=32. Significance levels : $*$: 10% $**$: 5% $***$: 1%. Standard errors are in parentheses. Estimates are obtained using STATA ver.9. [a]Simultaneous (interdependent) duration models are the models developed in this paper. [b]A univariate duration model is a conventional duration model with a single duration dependent variable. Therefore, in Model (4), the equation for the liberalizing and autocratizing durations were estimated separately, as in Alemán (2011).

the variables remain the same except the economic variables are not statistically significant anymore. This is consistent with the fact that Mauritius has exceptionally high GDP levels compared to other countries in Africa.[17]

The estimated coefficient α_1, the effect of the survival duration on the liberalization duration, is now -0.959 and the estimated α_2, the effect of liberalization on the survival becomes 0.398, both statistically significant at the 1% level. These are still substantial magnitudes, suggesting that a year increase in survival decreases the transition duration by about 2.6 years, and an increase in the transition period increases the survival by 1.5 years.

Overall, I am inclined to keep this case in the data, because (1) the fact that Mauritius had never had the liberalization period is controlled by the *not independent* variable, and (2) theoretically, its exceptional economic performance (at least among all the African states) is a predictor of its regime stability and it can also be controlled for by the economic variables included in the models. One has to remember, however, that the statistical significance of the economic variables, at least for now, seems to be driven heavily by this single case, and it is not appropriate to generalize the finding regarding economic development.

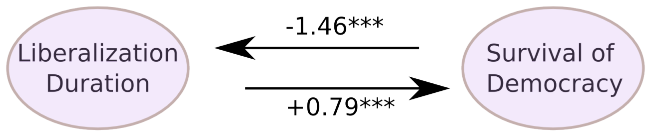
Simultaneity Biases: The Reason We Do Not Judge A Man "Not Moving" on a Treadmill

Earlier in the methodology section, I explained that the estimates from univariate analyses (i.e., explaining each duration by the other duration separately, as two distinct processes) potentially suffer from simultaneity biases. In fact, the biases turn out to be substantial in the case of the democratic transition and survival. In this section, I explain where in the results we can see the bias, and how dangerous it is to make conclusions about the relationship between the two durations by analyzing them in isolation.

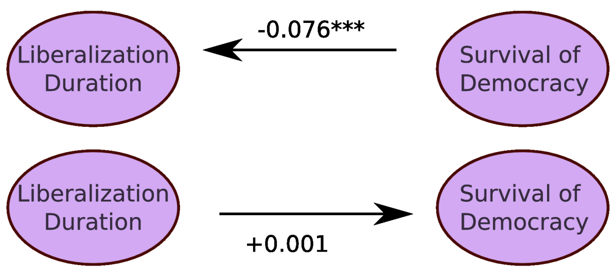
As I summarized above, the multivariate analysis with the new duration estimator reveals

[17]Its main industry is tourism.

that both durations have statistically significant effects on each other. More precisely, the liberalization transition has a positive effect on the survival of democracy, and the survival of democracy has a negative effect on the liberalization duration. Model (3), for example, demonstrates the following two-way causal relationships;



When these true causal relationships exist in both directions as in this case, what would happen if we estimate the effect of a duration on the other duration respectively for each duration? Model (4) shows us clearly what would happen. Recall that in Model (4), two processes are estimated separately. The model for democratization and survival each reveals;



First, the effect of liberalization on the survival of democracies is not statistically significant anymore. Furthermore, both magnitudes are much smaller than those from the bivariate estimation (Model (3)). The difference between the two models demonstrate simultaneity biases very clearly. In reality, there is a fairly strong negative effect of survival on the liberalization transition, and a weaker but still a solid effect of liberalization on survival.

However, when the two processes are artificially separated by model construction (Model (4)), the estimates can only reveal net effects. In this study, where the true signs of the two causal effects are opposite, the magnitude of the "causal" effects in the univariate model look much smaller than they really are, suffering from attenuation biases. This is similar to a "treadmill effect". Think of the effect of the liberalization duration as the speed of a man walking on a treadmill, and the survival duration as the speed or incline of the treadmill. The interdependent duration model reveals the fact that the man is walking very fast *and* the treadmill is also moving very fast, in the opposite direction to that of the man working on the treadmill. However, the univariate approach can only reveal the fact that overall the person on the treadmill is not moving very much–after all, he is remaining in the same spot in the gym. This is the confounded small effect we are seeing in the results from model (4). Therefore, making any conclusions from the univariate analysis would be equivalent to tell the person working on a treadmill that he is not doing any exercise because he is not moving anywhere. This is, indeed, a biased conclusion based on biased estimates.

2.7 Conclusion

2.7.1 Methodological Innovations

Politics generates interdependence across durations and duration interdependence across actors. There are many examples of this interdependence in salient areas of political science research. In order to analyze interdependent durations empirically, one has to make assumptions about the structure of dependence. And the structure of such dependence comes from our substantive theories. In many instances, the simultaneous equations framework provides a better match to substantive theories we posit. Therefore, I developeded a generalized parametric simultaneous equations model for interdependent duration processes and derived the corresponding full information maximum likelihood (FIML) estimator based on the Weibull distribution. Along the way, I also attempted to demonstrate why a copula-based estimator is unnecessary for our purposes, particularly given the added complications. I demonstrated with Monte Carlo experiments that the new interdependent duration estimator outperforms the alternatives.

2.7.2 Application to a Study of Democratization and Consolidation

The applicability of this estimator is plenty. One such example is the study of the determinants of democratic transitions and consolidation in Africa. The interdependence I uncovered in these durations is substantively important. The findings suggest that the duration of liberalization and the survival of democracies are interdependent, and the causal relationships between the two durations exist in both directions. It is indicative that the democratic oppositions do form anticipation about the future prospect of their anti-government movements, affecting the duration of the liberalization period, in a way the potential of a more durable democracy in the future shortens the duration of the democratic transition. At the same time, a longer transition duration positively influences the survival of a future democracy. This is consistent with the idea proposed by Almond and Verba, and Putnam. When the transition to a democracy is slower, there is a possibility that the cultures that appreciate democratic values can emerge in the society, contributing to the stability of a future democracy. Comparing the results from my new approach to the study of democratization and the traditional univariate approach (where the determinants of the emergence and survival of democracies are studied separately), it becomes evident that the traditional approach has been severely suffering from simultaneity biases. With the traditional approach, the effect of anticipation about the future regime outcome is very much underestimated, and the effects of the transition speed on the survival of democracies is not detected at all.

Chapter 3

Essay2
Right-Censoring in Interdependent Duration Models with the Systems of Duration Equation Modeling (SDEQ) Approach

Abstract

We say that an observation is *right-censored* when the end of the data-collection period comes before we observe the end of the duration of the observation. For example, in a study of democratization, the duration of continuing democracies at the end of the collected data is right-censored. When some observations are right-censored, estimation results from duration analyses can be misleading without a special treatment for it, because, for example, democracies that collapsed after 5 years, and democracies whose survival was observed *at least* for five years (but could have survived much longer after the data collection) are qualitatively different. Given the ubiquity of right-censoring in the social-science data, the basic interdependent duration model developed in Essay 1 is less useful without incorporating the censoring information in the likelihood. In this essay, I derive a likelihood function that accounts for right-censoring for the interdependent duration model developed in Essay 1. The primary methodological difficulty comes from the univariate and joint cumulative distribution functions (CDF) of the duration variables in the likelihood, for which we do know the exact expressions. In fact, many researchers believe that one has to "simulate" to estimate the model due to the complexity of the likelihood involving the CDF's of duration variables, and the simulation approach takes time and a substantial computational power. I overcome these issues and derive a simpler and exact expression of the likelihood function by applying the change-of-variables theorem multiple times. To estimate my model, one does not need to simulate, and the required computational power is very low.

I illustrate the use of the estimator with the new likelihood by applying it to the study of democratization in Africa (continued from Essay 1). The study of the democratic transition and consolidation is particularly an interesting application for this methodology, because among many topics in political science, both transitions and consolidation are extremely slow-moving phenomena. As a consequence, many survival data of democracies are right-censored. I find that, once I account for right-censoring, the signs of directed duration dependencies become opposite, and the effects of many more covariates become statistically significant, telling us much richer stories about democratization in Africa. I also compare the empirical results with those from univariate analyses conducted separately for the transition duration and the survival duration. These are the kinds of duration models that researchers would typically use, when they suspect dependencies between the two durations but do not use an interdependent (or multivariate) duration model. The right-censored observations are properly treated in these univariate regressions. Even then, the differences between the univariate and multivariate duration analyses are stark. The statistical significance, the sign of the effects (both for the duration dependency and covariates), and the magnitude are all very different between the two approaches, leading researchers to conclude differently. A set of Monte Carlo simulations to examine potential strengths and weaknesses of the likelihood and my Stata code is a task to be done in the near future.

3.1 Overview

In the previous essay, I developed an interdependent duration model, using a system of simultaneous equations. One of the fundamental methodological contributions was to develop a model with which one can estimate the existence and magnitude of interdependence between multiple durations, as well as the effects of structural covariates on each of the durations, within a single statistical framework. More importantly, my systems-of-duration-equations (SDEQ) modeling approach is a "substantive approach," which recovers the structure of dependency among multiple durations in a way that scholars should find useful. In other words, I directly *model* the structure of the dependency by expressing a duration as a function of the other durations, instead of treating the dependency as a "nuisance". I also compared my SDEQ modeling approach with a copula-based modeling approach for duration dependencies, which, instead, tends to be associated with the interpretation of dependency as a *nuisance* that biases the estimates of other covariates predicting durations.

However, each modeling approach has its own strengths and weaknesses. My previous work on SDEQ models provides only a basic estimation method to deal with multiple duration processes. For example, this estimator is useful as it is, if the data are *uncensored*. To clarify the terminology, we say an observation is *right-censored* when the end of data collection comes before we can see the end of the duration for the given observation. For example, if one is interested in the survival of democracies, the continuing democracies at the end of the data-collection period are right-censored (e.g., Alemán and Yang 2011; Svolik 2008). Right-censoring can also arise from legal constraints, as well as data-collection constraints. The duration of cabinet survival, for example, features two sources of right-censoring: first, cabinets that are still surviving at the end of the data collection period are right-censored observations; second, cabinets that lasted until the end of the constitutional inter-election period (CIEP) and thus legally "had to end" are also right-censored, because they could have presumably carried on governing but for the imposed cut-off (e.g., Laver and Shepsle 1996; Lupia and Strøm 1995; Warwick and Easton 1992; Alt and King 1994; Diermeier and Stevenson 1999). Social scientists are not always fortunate enough to have "low-maintenance" data, where all the relevant processes are observed from the beginning to the end without interruptions. Generally, the existence of right-censored observations can be, in fact, one of the reasons to use the duration approach, because a modification in likelihood functions in duration models can account for censoring (Box-Steffensmeier and Jones 2004).

Accommodation of right-censored observations is fairly simple to conceptualize and implement in single-duration models and interdependent duration models that are based on copulas. It is, on the contrary, not mathematically straightforward to account for right-censoring in the SDEQ approach. However, remember that the SDEQ approach has an advantage of modeling interdependence in a substantively meaningful way, compared to the copula approach. Therefore in this paper, I make a first attempt to develop a relatively simple way to evaluate an interdependent-duration likelihood function with a treatment that can account for right-censoring. Given the ubiquity of duration interdependence (see Essay 1) and right-censored observations in political science (Box-Steffensmeier and Jones 2004), the likelihood I derive here should be broadly applicable.

To understand both the source of difficulties and my strategy of deriving the likelihood, a mathematical explanation is necessary. Before going into details about the mathematical derivation of the likelihood, however, here is a brief summary of difficulties associated with this task. The difficulty of writing the likelihood function with right-censored data, particularly for models with interdependent durations, stems from the fact that the likelihood function involves the joint survivor function of random variable y's (durations). Equivalently, this can also be interpreted as the difficulty in deriving the joint cumulative density function (CDF) because the survivor function can be defined by the joint CDF. The relationship is $S(\mathbf{y}) = 1 - F(\mathbf{y})$, where $S(\cdot)$ is the joint survivor function and $F(\cdot)$ is the joint CDF. Deriving a joint survivor or cumulative density function involves multiple integrals with respect to all the duration variables. Multiple integrals become relevant only when we have right-censored data, because observations contribute information to the likelihood through the joint survivor function only when these observations are right-censored. In other words, we use information about event failures (represented by the PDF's of duration variables) when durations are completely known (uncensored), and we use information about the event survival (represented by the survivor functions) when observations are right-censored. A general likelihood function contains both components, with an indicator variable for observations that are censored.

This multiple integrals make estimation extremely burdensome. Hays (2009) and Franzese, Hays and Schaffer (2010) discuss this problem and suggest a possible estimation method using recursive importance sampling (RIS) and Bayesian Markov-Chain Monte-Carlo (MCMC). In this paper I take a different approach to simplify the likelihood. By using the change of variables theorem for multiple integrals, I change the main arguments of the joint PDF from y's to u's. By doing so, I obtain a likelihood with multiple integrals of u's joint density. Since u's are independent of each other, the joint PDF is in fact a simple product of univariate PDF's for all u's. Now each of the univariate integrals becomes independent in terms of calculating its value. Note that each integral of each univariate PDF is, by definition, a univariate CDF of a u. Thus, the entire multivariate integral term becomes a simple product of the univariate CDF of all the u's. We know the exact expression of each u, so we have the likelihood.

Again, the reason I make such efforts to derive a likelihood for the SDEQ approach (not using an off-the-shelf copula approach to incorporate censoring) is that I want to maintain the framework of SDEQ approach, in order to *model* dependencies structurally based on substantive theories. This way, I can take advantage of the SDEQ approach to modeling interdependent durations, and yet account for censoring.

In this essay, I illustrate the method in a study of the democratic transition and consolidation in Africa. One could evaluate the performance of this estimator by Monte Carlo simulations.

This paper is organized as follows. First, I review the set-up of the interdependent duration model developed in the previous work. Second, I present a general form of the likelihood function with right-censored cases. Third, I derive the joint survival function using the change of variables theorem (CVT) for multiple integrals. Fourth, I explain the data generating process and evaluate the performance of this estimator using Monte Carlo experiments. Finally, I estimate a simultaneous durations model of the democratic transition and consolidation of countries in Africa, and of government formation and survival in Western European democracies.

3.2 Review of the SEQ Interdependent Duration Models

In this section, I review the basic structure of the simultaneous equations model for interdependent duration processes developed in Essay 1. It should clarify how the full information maximum likelihood estimator was constructed and how the likelihood without right-censoring appears in general.

3.2.1 Linear Parameterization of Weibull Durations (The AFT Model)

The dependent variables of interest, $\mathbf{z}$, are D distinct duration processes that have Weibull distributions with two parameters.[1]

$$z_{id} \sim \text{Weibull}(\lambda_d, \theta_d), \tag{3.1}$$

where $i = \{1, \cdots, N\}$ denotes the observational-unit index and $d = \{1, \cdots, D\}$ denotes the duration index, implying that there are $N \times D$ observations in total. The notation λ is the shape parameter and θ is the scale parameter. These distributional parameters take common values across the N observational units; hence they have only one subscript that indicates duration process. The following density and distribution functions characterize the

[1] To distinguish an original Weibull random variable and a logged form of Weibull random variable (which has the type I extreme value distribution, or equivalently standard Gumbel distrbution) clearly, I use a symbol z for a Weibull variable, and y for a logged Weibull (type I EVD). Hence the relationship between them is $y = \ln(z)$.

univariate Weibull distribution;

$$\text{Weibull} \begin{cases} f(z_d) & = \frac{\lambda_d}{\theta_d}\left(\frac{z_d}{\theta_d}\right)^{\lambda_d - 1} e^{-(\frac{z_d}{\theta_d})^{\lambda_d}} \\ F(z_d) & = 1 - e^{-(\frac{z_d}{\theta_d})^{\lambda_d}} \end{cases} \tag{3.2}$$

A common way to parameterize a Weibull model of D interdependent durations is to log-linearize the model and obtain a log-linear system of D equations (Box-Steffensmeier and Jones 2004). It is also known that the logged Weibull variable turns out to be a standard Gumbel variable that is scaled by the shape parameter in the original Weibull distribution (λ_d). For example, in a univariate Weibull case, the log-linear form of a duration can be expressed as follows;

$$\begin{cases} y = \ln z = \ln\theta + \dfrac{1}{\lambda}u \\ \qquad\qquad = \mathbf{X}\boldsymbol{\beta} + \dfrac{1}{\lambda}u, \\ u \sim \text{Extreme Value I (Standard Gumbel), i.i.d.} \end{cases} \tag{3.3}$$

The following distribution and density functions characterize the standard Gumbel distribution;[2]

$$\text{Type-I Extreme Value (Standard Gumbel)} \begin{cases} f(u) = e^u e^{-e^u} \\ F(u) = 1 - e^{-e^u}. \end{cases} \tag{3.4}$$

[2]The standard Gumbel distribution is a special case of the type-I extreme value (minimum) distribution. The distribution and density functions of a general type-I extreme value (minimum) distribution are

$$\begin{cases} f(u) = \frac{1}{b} e^{\frac{u-a}{b}} e^{-e^{\frac{u-a}{b}}} \\ F(u) = 1 - e^{-e^{\frac{u}{b}}}, \end{cases}$$

where a is the location parameter and b is the scale parameter.

Hence the standard Gumbel distribution is a special case of the type-I extreme value distribution, where $a = 0$ and $b = 1$. A logged Weibull variable has the type-I extreme value distribution in general and only the scaling of the resulting extreme value variable varies depending on how one sets the scale parameter of the extreme value variable.

The second line of equation (3.3) shows how we could include covariates, by making the Weibull scale parameter, θ, a function of the covariates, $\theta = e^{\mathbf{X}\boldsymbol{\beta}}$. For further detail regarding the link between a Weibull and an extreme value distribution, see Appendix A.1.2.

3.2.2 A System of Multiple Duration Equations

A system of D distinct durations with N observational units in matrix notation is

$$\text{[Structural form] } \mathbf{y}_{(ND\times 1)} = \mathbf{A}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}. \tag{3.5}$$

The dependent variable, $y_{id} = \ln \mathbf{z}_{id}$, is a logged Weibull random variable. The vector $\mathbf{y}$ is a stack of D vectors, each of which contains N observational units.

$$\mathbf{y}_{(ND\times 1)} = \begin{pmatrix} \mathbf{y}_{.1} \\ \vdots \\ \mathbf{y}_{.D} \end{pmatrix}, \text{ where } \mathbf{y}_{.d(N\times 1)} = \begin{pmatrix} y_{1d} \\ \vdots \\ y_{Nd} \end{pmatrix}.$$

The matrix $\mathbf{A}$ is the coefficient matrix for the dependence. An element matrix $\boldsymbol{\alpha}^{d'}_{.d}$ contains coefficients representing the effects of the second duration d' on the first duration d. The diagonal elements $\boldsymbol{Sp}$'s in the $\mathbf{A}$ matrix are the matrices that capture the "spatial" dependency. This is the dependency among N observational units within each duration process. We call it "spatial" dependency for convenience, because the linear system captures the among-unit dependency using weights matrices just like in spatial contexts. Note that $\boldsymbol{Sp}_d = \mathbf{0}$ for all d when one assumes no among-unit dependency. Similarly $\boldsymbol{\alpha}^{d}_{.d'} = \mathbf{0}$ when one assumes no dependency between duration d and d'.

$$\mathbf{A}_{(ND\times ND)} = \begin{pmatrix} \boldsymbol{Sp}_1 & \boldsymbol{\alpha}^2_{.1} & \cdots & \boldsymbol{\alpha}^D_{.1} \\ \boldsymbol{\alpha}^1_{.2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\alpha}^D_{.D-1} \\ \boldsymbol{\alpha}^1_{.D} & \cdots & \boldsymbol{\alpha}^{D-1}_{.D} & \boldsymbol{Sp}_D \end{pmatrix},$$

where

$$\boldsymbol{\alpha}^d_{.d'(N\times N)} = \begin{pmatrix} \alpha^d_{.d'} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \alpha^d_{.d'} \end{pmatrix}, \boldsymbol{Sp}_{d(N\times N)} = \begin{pmatrix} 0 & \alpha_d^{(1,2)} & \cdots & \alpha_d^{(1,N)} \\ \alpha_d^{(2,1)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_d^{(N-1,N)} \\ \alpha_d^{(N,1)} & \cdots & \alpha_d^{(N,N-1)} & 0 \end{pmatrix}$$

The vector $\mathbf{x}$ denotes a set of covariates and the supscript indicates to which equation the covariate vector is specific. Each vector $\mathbf{x}$ contains K covariates with coefficients denoted by β. The subscript of $\mathbf{X}$, $.d$, indicates that these x's affect duration d, and the number of covariates, i.e., the number of elements in each $\mathbf{X}_{.d}$ is denoted K_d. The error term u_{id} in this structural form is i.i.d. with the extreme value minimum distribution. The error term is multiplied by λ_d^{-1}, which is the shape parameter of the original Weibull distribution and the value of λ is allowed to vary across duration processes.

$$\mathbf{X}_{(ND\times(K_0+\cdots+K_D))} = \begin{pmatrix} \mathbf{X}_{.1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{.2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}_{.TD} \end{pmatrix}, \text{ where } \mathbf{X}_{.d(N\times K_d)} = \begin{pmatrix} x^1_{1d} & \cdots & x^{K_d}_{1d} \\ \vdots & \ddots & \vdots \\ x^1_{Nd} & \cdots & x^{K_d}_{Nd} \end{pmatrix}$$

$$\boldsymbol{\beta}_{(K_0+\cdots+K_D\times 1)} = \left(\begin{array}{ccc|ccc|cc|ccc} \beta^1_{.1} & \cdots & \beta^{K_1}_{.1} & \beta^1_{.2} & \cdots & \beta^{K_2}_{.2} & \cdots & \cdots & \beta^1_{.D} & \cdots & \beta^{K_D}_{.D} \end{array} \right)';$$

$$\mathbf{L}_{(ND\times ND)} = \begin{pmatrix} \mathbf{L}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{L}_D \end{pmatrix}, \text{ where } \mathbf{L}_{d(N\times N)} = \begin{pmatrix} \frac{1}{\lambda_d} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{\lambda_d} \end{pmatrix};$$

$$\mathbf{u}_{(ND\times 1)} = \begin{pmatrix} u_{11} \\ \vdots \\ u_{ND} \end{pmatrix}.$$

The following reduced form can be derived from the structural form (3.5);

$$\text{[Reduced form] } \mathbf{y}_{(ND\times 1)} = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}-\mathbf{A})^{-1}\mathbf{L}\mathbf{u}. \tag{3.6}$$

3.2.3 Deriving the Likelihood via the Change of Variables Theorem for Density Functions

If we do not need to worry about right-censored observations, then the likelihood function takes the following form;

$$L = \prod_{i=1}^{N}\prod_{d=1}^{D} h(y_{11}, \cdots, y_{id}, \cdots, y_{ND}), \tag{3.7}$$

where $h(y_{11}, \cdots, y_{ND})$ is the joint pdf of the interdependent duration variables. The only task left to construct the likelihood function is to derive $h(\mathbf{y})$, the joint density of y's. We do not know the joint distribution of y's, but fortunately the change of variables theorem allows us to change the main argument from y's to u's. The advantage of this transformation is substantial, because u's are assumed to be independent (i.i.d.), and we know that the marginal of u has the type I extreme value distribution whose density function is well-defined. I use the change of variables theorem (for density functions) to derive the joint pdf of y's from the joint pdf of u's. By solving equation (3.6) for $\mathbf{u}$, we have

$$\mathbf{u}_{ND\times 1} = g^{-1}(\mathbf{y}) = \mathbf{L}^{-1}(\mathbf{I} - \mathbf{A})\mathbf{y} - \mathbf{L}^{-1}\mathbf{X}\boldsymbol{\beta}. \tag{3.8}$$

Johnston and DiNardo (1997) summarizes the theorem as follows;[3]

> *In the multivariate case* $\mathbf{u}$ *and* $\mathbf{y}$ *now indicate vectors of, say,* n *variables each. The multivariate extension of the previous results is*
>
> $$f(\mathbf{y}) = f(\mathbf{u}) \left| \frac{\partial \mathbf{u}}{\partial \mathbf{y}} \right| \tag{3.9}$$
>
> *where* $|\partial \mathbf{u}/\partial \mathbf{y}|$ *indicates the absolute value of the determinant formed from the matrix of partial derivatives,*
>
> $$\begin{pmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \cdots & \frac{\partial u_1}{\partial y_n} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \cdots & \frac{\partial u_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_n}{\partial y_1} & \frac{\partial u_n}{\partial y_2} & \cdots & \frac{\partial u_n}{\partial y_n} \end{pmatrix}$$
>
> *The absolute value of this determinant is known as the Jacobian of the transformation from* $\mathbf{u}$ *to* $\mathbf{y}$.

Going back to my duration model, the Jacobian of $\mathbf{u} = g^{-1}(\mathbf{y})$ is

$$\mathbf{J} = \left| \det \left(\frac{\partial \mathbf{u}}{\partial \mathbf{y}} \right) \right| = \left| det \begin{pmatrix} \frac{\partial g_{11}^{-1}(\mathbf{y})}{\partial y_{11}} & \cdots & \frac{\partial g_{11}^{-1}(\mathbf{y})}{\partial y_{ND}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{ND}^{-1}(\mathbf{y})}{\partial y_{11}} & \cdots & \frac{\partial g_{ND}^{-1}(\mathbf{y})}{\partial y_{ND}} \end{pmatrix} \right| . \tag{3.10}$$

---

[3] I have noticed that most textbooks call $|det(\frac{\partial \mathbf{u}}{\partial \mathbf{y}})|$ a Jacobian, but some call only $\frac{\partial \mathbf{u}}{\partial \mathbf{y}}$ a Jacobian. The bottom line is we need $|det(\frac{\partial \mathbf{u}}{\partial \mathbf{y}})|$.

If the inverse vector function, $(u_{11}, \cdots, u_{ND}) = g^{-1}(y_{11}, \cdots, y_{ND})$, exists for all $\mathbf{y} = (y_{11}, \cdots, y_{ND})$ such that $\mathbf{y} \in \{\mathbf{y} = g(\mathbf{u})\}$, the joint density of $\mathbf{Y} = g(\mathbf{U})$ is given by

$$
\begin{aligned}
h(y_{11}, \cdots, y_{ND}) &= \begin{cases} f(g_{11}^{-1}(y_{11}, \cdots, y_{ND}), \cdots, g_{ND}^{-1}(y_{11}, \cdots, y_{ND}))\mathbf{J} \\ 0, \text{ otherwise} \end{cases} \\
&= \begin{cases} f(u_{11}, \cdots, u_{ND})\mathbf{J} \\ 0, \text{ otherwise} \end{cases} \\
&= \begin{cases} f(u_{11})f(u_{12}) \cdots f(u_{ND})\mathbf{J} \\ 0, \text{ otherwise.} \end{cases}
\end{aligned}
\tag{3.11}
$$

The last line in equation (3.11) follows from the i.i.d. assumption of u, and each $f(u_{id})$ is the standard Gumbel pdf.

Therefore, if there is no censoring, the likelihood function is as follows;

$$
\begin{aligned}
L &\propto h(y_{11}, \cdots, y_{ND}) \\
&= \prod_{i=1}^{N} \prod_{d=1}^{D} f(u_{id})\mathbf{J}
\end{aligned}
\tag{3.12}
$$

Accordingly, the log-likelihood function is

$$
\ln L = \sum_{i=1}^{N} \sum_{d=1}^{D} \ln f(u_{id}) + \ln \mathbf{J} \tag{3.13}
$$

For example, if there are only two duration processes included in the model, the likelihood and log-likelihood functions look as follows.

$$\begin{aligned} L &= \prod_{i=1}^{N} \prod_{d=1}^{2} f(u_{id}) \mathbf{J} \\ &= \prod_{i=1}^{N} f(u_{i1}) f(u_{i2}) \mathbf{J} \\ &= \prod_{i=1}^{N} (e^{u_1 - e^{u_1}})(e^{u_2 - e^{u_2}}) \mathbf{J}, \end{aligned} \tag{3.14}$$

where

$$\mathbf{J} = \left| \det \begin{pmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} \end{pmatrix} \right| = \left| \det \begin{pmatrix} \lambda_1 & -\alpha_1 \lambda_1 \\ -\alpha_2 \lambda_2 & \lambda_2 \end{pmatrix} \right| = \lambda_1 \lambda_2 |1 - \alpha_1 \alpha_2|$$

and $u_1 = \lambda_1(y_1 - \alpha_1 y_2 - \mathbf{X}_1 \boldsymbol{\beta}_1)$. ($u_1$ and u_2 are symmetric.)

The log-likelihood function is;

$$\begin{aligned} \ln L &= \sum_{i=1}^{N} \sum_{d=1}^{2} \ln f(u_{id}) + \ln \mathbf{J} \\ &= \sum_{i=1}^{N} \{\ln f(u_{i1}) + \ln f(u_{i2}) + \ln \mathbf{J}\} \\ &= \sum_{i=1}^{N} \{(u_1 - e^{u_1}) + (u_2 - e^{u_2}) + \ln \mathbf{J}\}, \end{aligned} \tag{3.15}$$

3.3 Derivation of the Likelihood Function with Right-Censored Observations

If the joint survivor function $S(y_{11}, \cdots, y_{ND})$ is known, the likelihood function with right-censoring is

$$
\begin{aligned}
L &\propto \Big\{h(y_{11}, \cdots, y_{ND})\Big\}^{\delta_i} \Big\{S(y_{11}, \cdots, y_{ND})\Big\}^{1-\delta_i} \\
&= \prod_{i=1}^{N} \Big\{h(y_{i1}, \cdots, y_{iD})\Big\}^{\delta_i} \Big\{S(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i} \\
&= \prod_{i=1}^{N} \Big\{\prod_{d=1}^{D} f(g^{-1}(y_{id}))|\det(\mathbf{J})|\Big\}^{\delta_i} \Big\{S(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i} \\
&= \prod_{i=1}^{N} \Big\{\prod_{d=1}^{D} f(u_{id})|\det(\mathbf{J})|\Big\}^{\delta_i} \Big\{S(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i} \\
&= \prod_{i=1}^{N} \Big\{\prod_{d=1}^{D} f(u_{id})|\det(\mathbf{J})|\Big\}^{\delta_i} \Big\{1 - \sum_{d=1}^{D} F(y_{id}) + F(y_{i1}, \cdots, y_{iD})\Big\}^{1-\delta_i}
\end{aligned}
\tag{3.16}
$$

where δ_i is a not-censored indicator, taking the value 1 if observation i is not right-censored, and 0 if observation i is right-censored. Note that right-censored observations contribute to the likelihood only through the survivor function $S(\cdot)$ and not through the probability density $f(\cdot)$, which captures the failure information. This structure reflects the fact that, for right-censored observations, we know how long they have *survived* until now but we do not know when they will fail.

The survivor function can also be expressed with the joint CDF of the duration variables as $S(\mathbf{y}) = 1 - \sum_{d=1}^{D} F(y_d) + F(\mathbf{y})$ (Nelsen 2006). It turns out that, again, the change of variables theorem (applied to multiple integrals in this case) is useful to derive each of the univariate CDF for the duration variables y_d and the joint CDF $F(\mathbf{y}_{\cdot 1}, \cdots, \mathbf{y}_{\cdot D})$.

3.4 Deriving the Joint CDF by Change of Variables for Multiple Integrals

To the best of my knowledge, at least in political science, scholars have believed that it is extremely difficult to estimate any models that include likelihood functions written with joint CDF's of the dependent variables. For example, an earlier version of Franzese, Hays and Schaffer (2010) succinctly states the problem of potentially dealing with N-dimensional normal joint CDF (meaning dealing with N-variate integrals of normal PDFs) as follows;

> [t]his paper notes and explains some of the severe challenges posed by spatial interdependence in binary-outcome models and then follows recent spatial-econometric advances to suggest two simulation approaches for surmounting the analytically intractable and computationally intense estimation demands of these models, frequentist recursive-importance-sampling (RIS) or Bayesian MCMC. In brief, the complications arise because the endogenous spatial-lag implies the conditional independence that typically yields likelihoods for maximization that simply multiply N univariate distributions will not obtain. With interdependent observations, the likelihood is instead one N-variate joint distribution, and the one N-dimensional cumulative-normal in spatial probit is tremendously more intense to compute than the N cumulative standard-normal distributions of the common probit.

Another example is Hays (2009). This study defines the source of the estimation challenge more generally as a common problem for "simultaneous equation models with limited and qualitative dependent variables." Similarly to the spatial probit model developed in Franzese, Hays and Schaffer (2010), this study suggests a possible estimation method using recursive importance sampling (RIS), a simplified brute-force numerical-search approach for finding the maximum likelihood.

Here, I point out that, when the error terms of the structural form are assumed to be i.i.d., we can avoid such multiple integrals by using the change of variables theorem for multiple integrals. In the following, I demonstrate the mathematical procedure, and derive the likelihood function for an interdependent duration model with the SEQ approach accounting for right-censoring.

To recap the process up to the likelihood derivation, we start with duration dependent variables with certain probability distributions, and log-linearize the dependent variables to construct a linear parametric model of the covariates' effects on the durations. The structural form of the log-linearized duration equations has error terms that are assumed to be i.i.d.. Also importantly, we know the univariate PDF and CDF of the error term. For example, if the original duration variable is assumed to be Weibull distributed (as it is in this paper), each error term in th log-linearized form has the type I extreme value distribution (EVD). To construct the likelihood function of the logged dependent variables for uncensored observations, we need to derive the joint density of the dependent variables. This is done by the change-of-variables theorem for density functions, using the fact that each error term has a known distribution *and* these errors are independently distributed.

Once we start incorporating right-censoring, however, there is another function that we do not know. That is the joint CDF of y's in the following. This is the general form of the likelihood with censored observations (equivalent to equation (3.16)).

$$\begin{aligned} L &= \prod_{i=1}^{N} \Big\{ \prod_{d=1}^{D} f(u_{id}) |det(\mathbf{J})| \Big\}^{\delta_i} \Big\{ 1 - \sum_{d=1}^{D} F(y_{id}) + F(y_{i1}, \cdots, y_{iD}) \Big\}^{1-\delta_i} \\ &= \prod_{i=1}^{N} \Big\{ \prod_{d=1}^{D} f(u_{id}) |det(\mathbf{J})| \Big\}^{\delta_i} \\ &\times \Big\{ 1 - \sum_{d=1}^{D} \overbrace{F(y_{id})}^{F(y_{id})} + \overbrace{\Big(\int_{-\infty}^{y_{i1}} \cdots \int_{-\infty}^{y_{iD}} f(s_{i1}, \cdots, s_{iD}) ds_{i1} \cdots ds_{iD} \Big)}^{F(\mathbf{y})} \Big\}^{1-\delta_i}. \end{aligned} \tag{3.17}$$

In equation (3.17), there are two parts for which we do not know the exact expression: one is the expression for each univariate $F_Y(y_{id})$ and the other is the joint CDF of y, $F(\mathbf{y})$. n this section, I will first derive the joint cdf $F(\mathbf{y})$. As in the second line of equation (3.17), we can transform this using the definition of a joint CDF—the integral of the joint PDF of y's with respect to all the duration processes, $y_1, \cdots, y_D$.

The multiple integral makes estimation extremely burdensome. Since scholars believe that there is no simpler analytical expression for this term without the integrals, they have resorted to simulation approaches, which are often computationally quite burdensome and sometimes involve long waiting time for likelihoods to converge (Hays 2009).

Instead, one can eliminate the integrals, taking advantage of the i.i.d assumption we make for the structural-form error terms $\mathbf{u}$.

3.4.1 Change of Variables for Multiple Integrals

Applying the change of variables theorem to equation (3.17), we can obtain the joint PDF of y's for observation i, only as functions of u's.

$$
\begin{aligned}
F(y_{i1}, \cdots, y_{iD}) &= \int_{-\infty}^{y_{i1}} \cdots \int_{-\infty}^{y_{iD}} f_Y(s_{i1}, \cdots, s_{iD}) ds_{i1} \cdots ds_{iD} \\
&= \int_0^{u_{i1}} \cdots \int_0^{u_{iD}} f_u(t_{i1}, \cdots, t_{iD}) \mathbf{J} dt_{iD} \cdots dt_{i1} \qquad \text{by the CVT} \\
&= \int_0^{u_{i1}} \cdots \int_0^{u_{iD}} f_u(t_{i1}) \times \cdots \times f_u(t_{iD}) \mathbf{J} dt_{iD} \cdots d_{i1} \quad \text{by independence of } u\text{'s} \\
&= \left[\int_0^{u_{i1}} f(t_{i1}) dt_{i1} \right] \left[\int_0^{u_{i2}} f(t_{i2}) dt_{i2} \right] \cdots \left[\int_0^{u_{iD}} f(t_{iD}) dt_{iD} \right] \\
&= F_u(u_{i1}) \times \cdots \times F_u(u_{iD}) \mathbf{J} \qquad \text{by definition of a univariate CDF.}
\end{aligned}
\tag{3.18}
$$

where $\mathbf{J} = |\det(\frac{\partial \mathbf{u}}{\partial \mathbf{y}})|$ is the Jacobian of $\mathbf{u}$, which is the absolute value of the determinant of the $\frac{\partial \mathbf{u}}{\partial \mathbf{y}}$, and $\frac{\partial \mathbf{u}}{\partial \mathbf{y}}$ is a matrix of the derivatives of u's with respect to y's;

$$\frac{\partial \mathbf{u}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial f_u(u_1)}{\partial y_1} & \cdots & \frac{\partial f_u(u_1)}{\partial y_D} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_u(u_D)}{\partial y_1} & \cdots & \frac{\partial f_u(u_D)}{\partial y_D} \end{pmatrix}. \tag{3.19}$$

We know the exact expression of the above $F(\mathbf{y})$, because we know the CDF of a univariate u, $F_u(u_d) = 1 - e^{-e^u_d}$.

3.4.2 Change of Variables for Univariate Densities of Duration Variables

Recall equation (3.17).

$$L = \prod_{i=1}^{N} \Big\{ \prod_{d=1}^{D} f(u_{id}) |det(\mathbf{J})| \Big\}^{\delta_i} \Big\{ 1 - \sum_{d=1}^{D} \overbrace{F(y_{id})}^{F(y_{id})} + \overbrace{\Big(\int_{-\infty}^{y_{i1}} \cdots \int_{-\infty}^{y_{iD}} f(s_{i1}, \cdots, s_{iD}) ds_{i1} \cdots ds_{iD} \Big)}^{F(\mathbf{y})} \Big\}^{1-\delta_i}.$$

The only unknown part left is the exact expression of $F(y_{id})$. Again, since we know that the error term u_{id} has the type-I extreme value distribution, it will be useful to change the variable. From the expression in (3.3), one might be able to tell that, without covariates $(\mathbf{X}\boldsymbol{\beta})$, u is equivalent to y scaled by λ. Or on could also derive $F_Y(y_{id})$ more formally using the change of variables theorem again. Either way, we get the following expression.

$$F_Y(y_{id}) = F_U(u_{id}) \Big| \frac{\partial u_{id}}{\partial y_{id}} \Big| = F_U(u_{id}) \lambda_d. \tag{3.20}$$

3.4.3 The Likelihood Function

Therefore, using equation (3.20) and (3.18), we get the likelihood function accounting for right-censoring as follows;

$$\boxed{L = \prod_{i=1}^{N} \left[\left\{ \prod_{d=1}^{D} f_u(u_{id})\mathbf{J} \right\}^{\delta_i} \left\{ 1 - \sum_{d=1}^{D} F_u(u_{id}) \left| \frac{\partial u_{id}}{\partial y_{id}} \right| + \prod_{d=1}^{D} F_u(u_{id})\mathbf{J} \right\}^{1-\delta_i} \right]} \tag{3.21}$$

Accordingly, the log=likelihood function is;

$$\begin{aligned} \ln L = \sum_{i=1}^{N} \Big[\delta_i \Big\{ \sum_{d=1}^{D} \ln f_u(u_{id}) + \ln \mathbf{J} \Big\} \\ + (1-\delta_i) \ln \Big\{ 1 - \sum_{d=1}^{D} \Big(F_u(u_{id}) \Big| \frac{\partial u_{id}}{\partial y_{id}} \Big| \Big) + \prod_{d=1}^{D} F_u(u_{id})\mathbf{J} \Big\} \Big] \end{aligned} \tag{3.22}$$

For example, if a model includes two duration processes and the structural errors are assumed to be type-I EVD, the log-likelihood is;

$$\begin{aligned} \ln L = & \sum_{i=1}^{N} \Big[\delta_i \Big\{ (u_{i1} - e^{u_{i1}}) + (u_{i2} - e^{u_{i2}}) + \ln(\lambda_1 \lambda_2 |1 - \alpha_1 \alpha_2|) \Big\} \\ & +(1-\delta_i) \ln \Big\{ 1 - \lambda_1 (1 - e^{-e^{u_{i1}}}) - \lambda_2 (1 - e^{-e^{u_{i2}}}) \\ & +(1 - e^{-e^{u_{i1}}})(1 - e^{-e^{u_{i2}}}) \lambda_1 \lambda_2 |1 - \alpha_1 \alpha_2| \Big\} \Big] \end{aligned} \tag{3.23}$$

3.5 Application: Democratic Transitions and Survival in Africa

I continue to use the example of democratic transitions and survival in Africa to illustrate the methodological findings in this essay. To repeat my research interest, the broad questions to be answered are: "what makes some countries more likely to undergo democratic transitions than others" and "what makes some countries remain democratic while others revert to dictatorships.” Specifically, I investigate the existence and strength of the typical structural (or country-specific) attributes in determining democracy levels, as well as the strength of dependence between the timing of liberalization and the survival of democracies.

The variables considered for this analysis are the same as Essay 1. However, unlike Essay 1, Mauritius is dropped from the sample for this study. Mauritius has an exceptionally long-lived democracy among all the African countries included in the study, and it has an exceptionally high GDP per capita level. In both Essay 1 and 2, I conducted analyses with and without Mauritius. In Essay 1, I reported the results with Mauritius, because the qualitative results were essentially equivalent between the two, except that the effect of GDP per capita was not statistically significant anymore without the country. However, once I introduce the new statistical model in Essay 2, accounting for right-censored observations, the qualitative results between the analyses with and without Mauritius differ tremendously, without showing any obvious patterns in differences. Therefore, I decided to exclude the observation from the data.

Since 16 out of the 32 sets of observations are right-censored in my data, the results show significant differences between the estimates from the model with and without an appropriate right-censoring treatment. Table 3.1 summarizes the estimation results.

3.5.1 Comparing the Models without (Model 1) and with (Model 2) Right-Censoring Information

It is quite obvious that it is always better to account for right-censoring statistically when one can, but to show the stark difference in estimation results, I listed both sets of results in Table 3.1. The first column shows the results without accounting for censoring, and the second column uses the new estimator accounting for censoring. The difference is evident.

First, many more suspected variables for the liberalization duration turn out to be statisti-

Table 3.1: Estimation Results for Duration Models of Liberalizing and Autocratizing Transitions, Accounting for Right-Censoring

	Interdependent Duration Models		**Exogenous Durations**
	Not Accounting for Censoring (1)	**Accounting for Censoring** **(2)**	**Accounting for Censoring** (3)
Liberalization duration (y_1)			
θ_1 (Scale parameter 1)			
Real GDP/cap	-0.153	-0.865**	-0.024
	(0.123)	(0.436)	(0.062)
(Real GDP/cap)^2	0.002	0.037**	0.0003
	(0.001)	(0.018)	(0.001)
Commonwealth	-0.065	1.621**	0.038
	(0.682)	(0.653)	(0.325)
Urban pop rate	0.339	-0.222	0.146
	(0.289)	(0.270)	(0.146)
Muslim pop rate	-0.158*	0.305**	-0.014
	(0.085)	(0.135)	(0.044)*
Military regime	0.894	3.850***	0.466
	(0.584)	(0.687)	(0.287)
Fuel export rate	-0.118	-0.281**	-0.024
	(0.110)	(0.120)	(0.055)
Intercept	8.244***	-1.337	3.003***
	(1.444)	(2.212)	(0.780)
α_1 Dependency 1			
Survival	-2.544***	3.441***	-0.027
	(0.014)	(0.000)	(0.046)
λ_1^{-1} (Shape parameter 1)			
Constant	0.745***	0.892***	0.638***
	(0.109)	(0.156)	(0.106)
Survival duration of democracies (y_2)			
θ_2 (Scale parameter 2)			
Real GDP/cap	-0.003	-0.029	0.077
	(0.013)	(0.132)	(0.053)
Commonwealth	-0.018	-0.418	0.084
	(0.231)	(0.747)	(0.262)
Urban pop rate	-0.047	0.151	0.043
	(0.121)	(0.362)	(0.136)
Muslim pop rate	-0.013	0.022	0.021
	(0.036)	(0.102)	(0.037)
Military regime before	-0.503*	0.990	-1.087***
	(0.270)	(1.019)	(0.411)
Not independent before	3.628***	-9.922***	1.147**
	(0.352)	(1.058)	(0.456)
Presidentialism	-0.161	0.067	0.054
	(0.255)	(0.880)	(0.295)
Intercept	-1.493***	11.791***	-0.401
	(0.395)	(1.493)	(0.765)
α_2 Dependency 2			
Liberalization	1.204***	-3.503***	0.087***
	(0.000)	(0.000)	(0.020)
λ_2^{-1} (Shape parameter 2)			
Constant	1.947***	0.930***	0.379***
	(0.289)	(0.066)	(0.000)
N	31	31	31
Log-likelihood	-43.87	6.14	-38.96/-16.77

Significance levels: $*$: 10% $**$: 5% $***$: 1%. All models estimated without Mauritius.

cally significant when the likelihood incorporates the censoring information. The economic variable has a negative effect on the liberalization duration: transitions are faster when the

GDP per-capita levels are higher. The quadratic term of the GDP variable suggests that the marginal decrease in the transition speed decreases as the economic development level increases. On the other hand, the economic variable is not significant for the survival duration in any models estimated in this study, suggesting that the economic development level has nothing to do with the survival of democracies. The urban population rate, which could be another indicator of development, did not turn out to be significant either. Overall, this finding is somewhat consistent with the endogenous democratization literature (Boix 2003), but it conflicts with the findings in Przeworski et al. (2000). In fact, Przeworski et al. (2000) predict and find nearly opposite results: the development level has nothing to do with the timing of transition, but it positively affects consolidation.

Some variables become statistically significant only when we account for censoring. For example, Commonwealth countries seem to experience longer transition durations than other countries. Military dictatorships last longer than other types of dictatorships such as monarchic or civilian dictatorships. Fuel export rates is now statistically significant and makes the liberalization faster. This conflicts with the typical resource curse associated with slower democratization (or continuing autocracies). This might be when one needs to look more into the distributional aspects of economic variables and not only the development "levels".

What is substantively more important is that the sign of estimates flip for some important variables. Not surprisingly, these three are all directly related to durations. First, once we account for censoring, we find that countries that democratized immediately after their independence, on average, experience shorter democracies. This might reflect the institutional preconditionist argument that democracies that were born under premature institutional conditions do not last long (e.g., Huntington 1991). Both of the duration dependencies also have opposite signs now. Once we account for censoring, the survival of democracies now has a positive effect on the liberalization duration, supporting the argument about elites' strategic decision-making. The authoritarian elites are more willing to democratize precisely when they expect a short-lived democracy in the future, or in other words, they attempt to prolong their tenure in office when they anticipate that the emerging democracy is a stable one. This is somewhat consistent with the formal-theoretic finding in Acemoglu and Robinson (2006*a*). It is also consistent with the broad strategic argument maintained in Fearon (1998), where the author argues that political actors negotiate harder on international agreements exactly when they anticipate the higher level of enforcement in the future. The other causal effect—the effect of liberalization on the survival of democracy—turns out to be negative suggesting that slower transitions tend to lead to shorter democracies. In the introduction, I focused on the institutional preconditionist argument, suspecting the positive effect of the liberalization duration on the survival duration; however, it seems to be the case that the authoritarian regime has negative legacies on the stability of future democracies.[4]

The lesson is clear: with a large proportion of the cases censored in our data set, models with and without explicit accommodation of censoring produce dramatically different empirical results.

3.5.2 Comparing the Interdependent (Model 2) and Univariate (Model 3) Duration Models

However, now that an interdependent duration model can account for right-censoring, it is more meaningful to compare the estimation results from this model to those from the existing univariate duration models (Model (2) vs. Model (3)). The univariate approach estimates the effects of structural variables (e.g., GDP per capita etc.) on each of the two durations separately. For the *liberalization* equation, the *survival* duration is simply added as one of the covariates in the right-hand side, and vice versa.

The differences in the two sets of results clearly demonstrate how misleading it would be to

[4]This also leads me to a different research question: do repeated but interrupted democratization experiences within a country positively affect the stability of its future democracies? Are durations of multiple repeated democratizations correlated?

use the univariate approach when there are two-way causal relationships between the durations of liberalization and survival. We can see substantial differences both in the estimates of the structural covariates (e.g., GDP per capita etc.) *and* the duration dependencies. Moreover, the differences appear in the statistical significance, the sign *and* the magnitude of estimates.

It is worth highlighting the difference in duration dependencies. In short, the estimated significance changes and the signs flip. Suppose that a researcher suspects the dependency between the two durations but he/she explains each duration separately as in Model (3). From this univariate analyses, one would conclude that there is no significant effect of the survival of democracies on the duration of liberalization (-0.027 with p-value 0.557), when in fact there is a strong positive effect (3.441 with p-value 0.000). Also, one would conclude that there is a positive causal effect of the liberalization duration on the survival of democracies (0.087 with p-value 0.000), when in fact, the causal effect is negative and the magnitude is much larger (-3.503 with p-value 0.000). With the univariate analyses, we would conclude that there is no strategic causal effect of the survival on transitions, and slower transitions lead to longer-lasting democracies. The world would look quite different than what we see from the interdependent duration model, Model (2).

3.6 Conclusion

This essay extended the basic interdependent duration model developed in Essay 1, such that the likelihood function accounts for right-censored observations. Since the new likelihood function now involves not only the joint PDF but also the joint CDF of the duration variables, it manifests its own mathematical challenges. I overcame this problem by applying the change of variables theorem three times, and eventually derived the exact likelihood function without leaving in multivariate integrals. This makes it easier for researchers to estimate without concerns about computational burdens, as one would have to in a simulation approach. I derived the likelihood function also without having to approximate it in any way.

With this estimator, applied researchers can account for statistically both (i) the effects of structural covariates on each of the multiple duration, as well as (ii) the strength of the directed causal relationships among the duration outcomes, even when some of the durations are right-censored. This innovation makes the model much more useful for social science studies, because right-censored durations are very common in our data.

Researchers should account for right-censoring in their duration analyses not only because "they can" with a new statistical tool. I demonstrated, with the real-world data on democratization in Africa, that the qualitative (or substantive) differences in estimation results are non-negligible.

Comparing the models with and without the right-censoring treatment, I showed that the signs of duration dependencies can become opposite, providing evidence for different theoretical arguments, and also found that the statistical significance of other covariates can be much more enhanced by taking into account the censoring information.

Substantively, I conclude that the duration of liberalization and the survival of democracies are dependent. It is suggestive that the authoritarian elites anticipate the outcome of liberalization, and attempt to make the transition slower, when they expect that the emerged democracy would last longer. Or in other words, the elites are more willing to democratize, precisely when they anticipate a short-lived (or a reversible) democracy in the future, as Acemoglu and Robinson (2006*a*) describe in their formal analysis. This mechanism is also similar to one of the formal-model findings in the seminal bargaining study by Fearon (1998): when political actors anticipate that there would be a high enforcement level of an international agreement, they negotiate harder in the earlier negotiation phase. At the same time, I find that when the liberalization process takes longer for any reasons, it can shorten the survival of the emerged democracy. This finding is not consistent with the typical (institutional) preconditionist arguments (Moore 1966; **?**). It also conflicts with the suggested positive effect of the civic-culture development on stable democracies (Almond and Verba 1965; Putnam 2002). This negative causal effect of the liberalization duration on the democratic survival might be revealing the legacy effect of authoritarian regimes (Grzymala-Busse 2006).

Chapter 4

Essay3
Spatial Econometric Approach to Coevolution: The Diffusion and Reinforcement of Political Regimes

Abstract

What determines countries' democracy or autocracy levels? This essay provides unified theoretical and empirical models of democracy, with a greater emphasis on *regime reinforcement* across countries and over time. As found in existing studies, I recognize that democracy levels depend on countries' own attributes, and also on other countries' regimes by learning mechanisms at the civic and elite levels. However, in my view, countries do not always "learn to change" and are not always "open to change" their regimes: they can also try to maintain their current democracy/autocracy levels, for better or worse. I introduce an idea of the regime-support/approval dependency, defined by the similarity in states' current democracy levels. I demonstrate, both theoretically and empirically (with time-series cross-sectional data for 1950-2000), that states with similar regimes indeed select themselves into regime-support networks, and influence each other's regimes, reinforcing them over time. By simulations, I also show that political regimes can globally diverge toward multiple clusterings—as we observe democracy and autocracy "clubs" in the real world—only when the reinforcement (or self-selection) dynamic is introduced in the model, suggesting that existing diffusion theories are not sufficient representations of democratization. I conclude this essay with discussions on the useful integration of the spatial econometric and social network analysis (SNA) approaches, in order to deal with more complex diffusion mechanisms of political phenomena. I present a research note for modeling history dependence (including path dependence) in democratization, using a hybrid spatial-SNA statistical model.

4.1 Overview

What determines countries' democracy or autocracy levels? To answer this long-standing question, both qualitative and quantitative studies have provided various theories explaining political regimes. This essay provides unified theoretical and empirical models of democracy, with a greater emphasis on the *regime reinforcement* dynamic over time across countries.

The earlier literature, represented by the seminal work by Przeworski et al. (2000) for instance, provided evidence to support the significant association between regime types and various country-specific attributes. These earlier studies theorize that country-specific characteristics such as the country's GDP level, GDP growth, and the surrounding war-and-peace status predict its democracy level. Another literature, led by O'Loughlin et al. (1998) for instance, emphasized international determinants of democracy levels. Mechanisms aside, these studies have revealed that similar political regimes tend to cluster both temporally and geographically. This is when scholars started to ask what determines the seeming clustering of regimes, or what determines the local convergence toward the autocracy and democracy "clubs", and this is where the theory of contagious political regimes emerged. They theorized that political regimes might be "contagious" across countries just as diseases and habits can travel across individuals who are well connected. Their primary interest was the association between countries' *geographical* proximity and the magnitude of clustering observed in political regimes.

Later, Beck, Gleditsch and Beardsley (2006) argued that "space is more than geography." In their view, geographical proximity is just one of many—perhaps the most obvious—regime-dependence patterns that can facilitate contagion of political regimes. To prove their point, they incorporated the concept of economic dependency among countries (measured by trade volumes) and showed that countries influence political regimes in other countries more significantly when they are economically more dependent, and when they are geographically close. Goodliffe and Hawkins (2011) took a step further and considered three non-geographical dependency networks (trade, alliance, and international organization (IO) partners) through which democracy and autocracy might spread across countries. Both of these studies offer insightful theories about the democratic (or authoritarian) diffusion; and yet, in my view, even their definitions of space is limited.

In this essay, I re-examine mechanisms of regime diffusion by incorporating yet another type of space, wherein countries' distance affects their inter-dependence. I first argue, in ways consistent with the existing concept of diffusion, that countries that are more strongly dependent on each other influence each other's regime more significantly. The spaces that define the state dependency in my study are geographical proximity and economic dependency through trade. However, the main claim that makes my diffusion theory unique is that states can also strengthen other states' democracy/autocracy levels, by either explicit economic support or an implicit political-ideological support. Because of the other states' support, a country can potentially strengthen—or at least maintain—its democracy (or autocracy) level over time. For a country that is already relatively democratic, reinforcing its level of democracy would mean to move further toward a higher democracy score, and for a country that is now relatively authoritarian, reinforcing its regime would mean to deepen its autocratic reign. An important question is, when or under what conditions would a state support others' regimes? In other words, on what dependency "*space*" does regime reinforcement operate across countries? In this essay, I argue that states actively support or at least passively approve of each other's regimes when their regimes are more similar than dissimilar. Hence the pathways of such influence should be defined by similarity in their democracy levels from the previous time period.

Why do I treat this particular type of dependency—regime-similarity network—differently from the ones used in the conventional diffusion studies such as geographical connections, economic, alliance and IO partners? The simplest answer to this question is that the construction of the two types of spaces is different. The former, the regime-similarity network, emerges from *homophily*, while the latter does not. Homophily is a phenomenon in which individual units form connections with similar others. McPherson, Smith-Lovin and Cook (2001) extensively review studies that demonstrate homophilic tie formation among individuals, which include situations where one becomes friends with others due to their similar behavioral habits (e.g., smoking, preferring to work in a certain industry etc.). The formation of regime-reinforcement network is a homophilic dynamic in which states form (mostly) ties with others when their regimes are similar. Networks defined by geographical proximity, on the contrary, are fixed by border placement regardless of countries' preferences. Economic networks are defined by numerous factors, only some of which reflect logic state similarity in any aspect. For example, trade networks depend largely on endowments and the logic of supply and demand for certain goods and services, geographical distance, and historical incidents that facilitate interstate transactions such as colonialism. None of these are purely correlated with the similarity in countries' political, economic and social aspects. Table 4.1 compares the two types of dependency spaces used in the traditional diffusion studies and my study.

Table 4.1: Characteristics of the Two Types of Dependency Spaces

Literature	Dependency networks defined by...	**Networks emerging by homophily?**
Traditional diffusion	Geographical proximity, Economic partners	**No**
My theory	Geographical proximity, Economic partners Regime similarity	**No** **Yes**

To understand the true importance of my regime-reinforcement mechanism, however, it is necessary to consider the long-run consequence of incorporating (or not incorporating) the homophily-based dependency network. The key dynamic to understand this long-run consequence is "*network-behavior coevolution*." The homophlic tie formation between similar regimes recalls the "one flap of a butterfly's wings" in a famous story of chaos theory: a butterfly's flap of wings at one point can change the course of the universe to the extent that it can cause a tornado on the other side of the globe. The moral of the story is: a small event at one point in a non-linear system can cause a significant differences to the course of the system evolution later. This sensitivity issue to initial (past) conditions is the heart of chaos or complex systems theory.[1] Why is this relevant to my theory of regime-reinforcement? It is because the reinforcement dynamic generates non-linearity in the theoretical model I construct. The non-linearity emerges from the following inter-temporal and inter-spatial loop, called *network-behavior co-evolution*. Countries with similar regimes form stronger dependency ties by homophily (see Figure 4.1).[2] In the next time period, a state's regime influences that of other states' by the supporting and approving mechanisms (see Figure 4.2). The influence is stronger when the tie between a given pair of countries is stronger. These new levels of democracy, in turn, regimes define the regime-similarity network of the next time period (see Figure 4.3). Hence, the latent connectivity across countries ("network") and their democracy level ("state behavior") co-evolve over time. This is the network-behavior coevolution, which induces non-linearity in the system of 170 plus countries. I do not expect that regime-reinforcement among states would have a dramatic effect on the level of democracy within a single time period, after controlling for the effects of state-level characteristics (the GDP level etc.) and the effects of others' regimes through conventional dependency paths. After all, the support of other regimes is normally subtle and nuanced. However, this seemingly small dynamic of states' self-selection into the reinforcement network operates just like a flap of a butterfly's wings.[3] As long as the reinforcement dynamic exists,

[1]A mathematician and meteorologist, Edward Lorenz, used an example of a butterfly's flap of wings to describe the initial condition sensitivity of complex non-linear systems. He used the example in his talk "Predictability: Does the Flap of a Butterfly's Wings in Brazil set off a Tornado in Texas?" at the American Association for the Advancement of Science (1972). (Source: `http://www.cmp.caltech.edu/~mcc/chaos_new/Lorenz.html`)

[2]In the figures, node color designates democracy level, and edge width signifies the strength of ties.

[3]Even though I will elaborate on my empirical strategy later, I would like to note here that the estimated coefficient in a typical regression table is an immediate short-term reinforcement effect. I expect the number to be fairly small. However, what matters most is whether the reinforcement factor turns out to be statistically significant or not; i.e., whether such a dynamic exists in the system. Does the regime-reinforcement butterfly exist in our system of 170 plus countries?

or as long as it is "statistically significant" in statisticians' language, the system based on my theory of democratization demonstrates a significantly different story in the long run from what we already know based on existing studies. The difference is mainly its sensitivity to the initial world-wide distribution of democracies. I will come back to this point when I discuss my simulation studies.

Figure 4.1: Homophilic Tie Formation by Regime Similarity

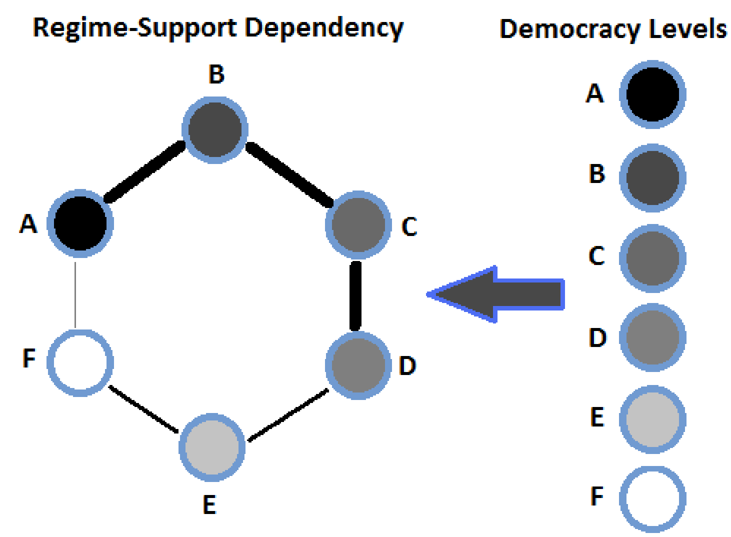


Figure 4.2: Regime-Reinforcing Influence (Contagion) Through a Regime-Support Network

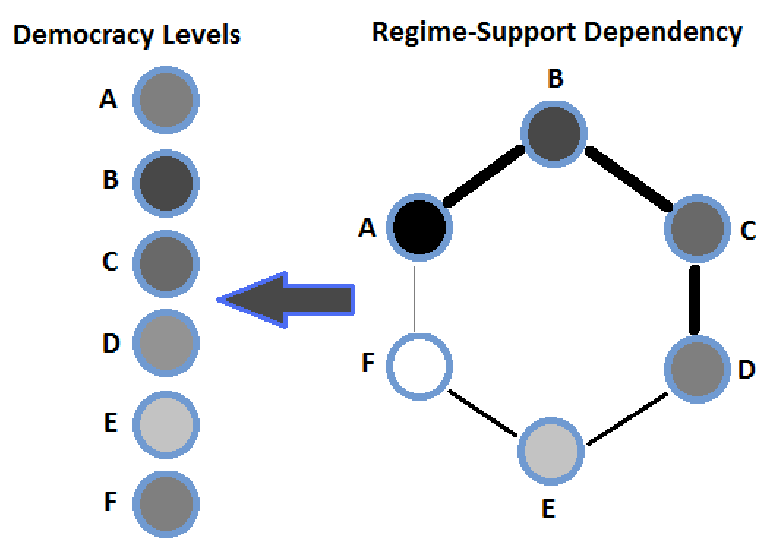
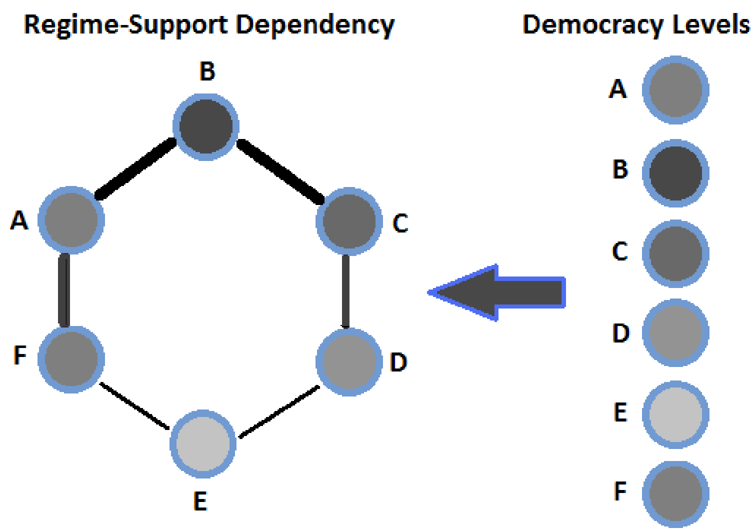


Figure 4.3: Homophilic Tie Formation by Regime Similarity



How does the regime-reinforcement theory fit in a fuller context of modeling democracy levels? To repeat my research question, it is "what determines countries' democracy or autocracy levels?" The main dependent variable is the level of democracy, as measured by the Unified Democracy Scores (Pemstein, Meserve and Melton 2010). While maintaining the emphasis on the reinforcing mechanism of political regimes across countries, my predictive model for the level of democracy also incorporates important existing theories. I argue that there co-exist three mechanisms that determine the level of democracy. First, as the earliest democratization literature suggests (e.g., Przeworski et al. 2000), the level of democracy largely depends on a country's own characteristics. I include several country-specific variables that represent states' economic indicators, colonial history, urbanization indicator, resource richness, and religious profile. The second is the diffusion process, in which a country can influence other's regimes through pre-defined dependency networks. This part of the model resembles the concept of conventional diffusion studies.[4] Finally, the third part of my model

[4] I incorporate the substantive theory of existing diffusion studies, but I employ a spatial econometric model, which is more intuitive than the method used in Goodliffe and Hawkins (2011) and also an improved version of a spatial-lag model compared to Beck, Gleditsch and Beardsley (2006). In short, spatial lags in my model do not need to be temporally lagged, as was the case in Beck, Gleditsch and Beardsley (2006). This means that my model can allow for instantaneous regime contagion within the same time period–within a year in the case these democratization studies, including mine.

reflects the regime reinforcement mechanism, wherein there is stronger diffusion of democracy or autocracy among countries that have more similar democracy scores.

Do data support my theory? And how can one statistically distinguish the three mechanisms built in to the theoretical model? The crucial methodological issue here is how to distinguish statistically between the three mechanisms determining the level of democracy. In the first mechanism, which I call *common exposure*, the level of democracy is a function of country-specific attributes such as the economic development level. The coefficients associated with these terms represent the extent to which a country's democracy level changes by its exposure to shocks in these country-specific covariates. The interpretation of these estimates will be the same as those in "standard linear regression models". In the second mechanism, which I refer to as *contagion*, a country's regime influences others' through pre-defined dependency networks such as geographical and economic connectivities. The estimates associated with these terms indicate to what extent contagion of regimes occurs through each of the given pathways. Hence the actual effects (or influences) of other countries' regimes would be a product of the (i) estimated coefficient, (2) the other countries' democracy scores and (3) spatial weights associated with all the possible pairs of countries, which measure how strongly countries are connected through the given network. The third mechanism is similar to the second, in that the term includes (1) a coefficient indicating how important the regime-support network is, (2) other countries' democracy scores, and (3) spatial weights for all the pairs of countries, which measure how similar regimes are for all the possible pairs. However, in the third mechanism, since the spatial weights at one time point are defined by regime scores of the previous time period, the reduced-form cross-sectional regression equation cannot be independent of the past time periods anymore. Unifying the three mechanisms requires a special treatment—in fact a new spatial model. In short, the inclusion of the third mechanism allows us to evaluate whether countries self-select themselves into networks of similar selves (in the dimension of democracy levels), and whether that network in turn becomes an important pathway of the interstate regime influence in the future. In the rest of the essay, I will use "*selection*" and "reinforcement" interchangeably to refer to this third mechanism.[5]

To estimate the magnitude of the three mechanisms—*common exposure*, *contagion* and *selection*—separately, I introduce a multiparametric spatiotemporal autoregressive (MSTAR) model, a new spatial-lag model (Hays, Kachi and Franzese 2010). The methodological innovations of this model are two-fold. First and most importantly, the MSTAR model is one of the first that can distinguish *selection* from *contagion*. Obviously, this methodological innovation was necessary for me to test for my hypothesis about regime reinforcement. Second, even only for estimating the second mechanism, contagion, an MSTAR model is more preferable to than the one presented in Beck, Gleditsch and Beardsley (2006).[6] To be clear, Beck, Gleditsch and Beardsley (2006) study the diffusion of democracy for only a single cross-section. Therefore the model used for the study of democracy in their paper is not comparable to what I present here for time-series cross sectional (TSCS) data. However, they also discuss and apply a spatial lag model for TSCS data to the study of trade flows. Since this model does not allow for selection, the only comparable part to MSTAR is the conventional contagion term. The difference is that their model allows for the influence of the other units *only from the previous time period*, while our MSTAR model allows for the possibility that contagion occurs simultaneously within a single time period. Permitting only a time lag of others' influence can be acceptable in some contexts, but in my view, it is a limitation to our empirical studies most of the time.

In my empirical analysis, I find that a country's democracy level indeed influences others'

[5]Both selection and reinforcement belong to the same loop associated with the third mechanism. The only difference is that the idea of selection emphasizes the endogenous network formation among countries that have similar regimes scores, and reinforcement emphasizes the resulting democracy levels partially determined by the homophilic network.

[6]As I mentioned earlier, the main focus of Goodliffe and Hawkins (2011) is also the diffusion of democracy, and the dynamic they theorize is exactly what a spatial model could implement in a relatively simple manner. However, the authors somehow decided to execute it in an OLS model with a creative dependency networks indicator and geographical proximity as covariates. To distinguish the democratic diffusion from self-selection (i.e. to eliminate the reverse causality from democracy levels to dependency networks), the author had to take a one-year time lag for all the diffusion-related covariates.

democracy levels through the regime-reinforcement (regime-support) network. This finding is quite noteworthy, given that this is after controlling for various country-specific covariates and the effects of others' regimes through two (non-homophilic) dependency networks (geographical and economic connections). The finding implies that countries do form latent support networks with other countries of similar regimes, and they do influence each other's regimes through this pathway. Even though this analysis is not an explicit test for local and global convergences of democracy levels, it is highly suggestive that the reinforcement mechanism is contributing to the local convergence of regimes that we observe in the world today—a stable group of democracies and also a stable (or stagnating) group of authoritarian regimes.[7]

As a way to conclude this essay and to motivate the next iteration of the diffusion study, I conduct two sets of simulations. They demonstrate the long-run effects of the reinforcement dynamic. As I mentioned earlier, the estimates reported in the regression table gauge the magnitude of the effect within a single time period, and these look quite small—just "a flap of a butterfly's wings". However, the simulation analyses show more dramatic differences between a model with and without the reinforcement (or self-selection) mechanism. First, I use the same regime data to show the long-run trajectories of all the countries' democracy scores after a democratic shock in China. I compare the trajectories using models with and without the selection component. These simulations provide us with a better idea of what long-run differences we might see between the two distinct systems; in other words, this is the difference that we would fail to capture, if we assume away the reinforcement dynamic by omitting the term, as conventional diffusion studies do. The second set of simulations has a slightly different focus. I compare two distinct systems: one with and the other without

[7]The "Arab Spring", a wave of protests and revolutions that started in the Middle East and North Africa the end of 2010, might be an exogenous shock that potentially changes the course of the regime distribution in the world, loosening the ties among the long-lasting authoritarian bloc in the region. The impact of this effect will also depend on what potential democratic (or semi-democratic) elections after these uprisings produces in each country. I think that we have to wait for a little longer to understand the true nature of this series of events.

the reinforcement dynamic. This time, however, I compare them in terms of how they evolve over time when we give three different sets of starting values—the same three sets to each of the two systems. The finding is quite astonishing. In the model without reinforcement, outcomes of all the units converge, and they converge to exactly the same level, while in the reinforcement model, one demonstrates a global convergence and the other two demonstrate global divergences, each to different levels. Even more interestingly, these two diverging cases also demonstrate two clusterings. The implication of this result is particularly noteworthy, because if I apply this finding to the context of democratization, the conventional spatial approach (i.e., without selection) would be assuming, just by the choice of the model, that all the states' democracy levels should converge, and would converge to exactly the same level of democracy at some point in the future, *regardless of* the distribution and topology of surrounding political regimes; i.e., the initial democracy profile of the world is irrelevant to a country's regime trajectory in the long run. This is a very unlikely scenario in the real world.

This essay has two appendices. Both briefly discuss topics in the intersection of spatial analysis and social network analysis (SNA). As Hays, Kachi and Franzese (2010, pg.407) state:

> [d]espite the obvious conceptual, theoretical, and substantive overlap between spatial and networks interdependence, spatial econometric and network analytic research have developed largely in isolation.

In Appendix A.3.1, I suggest uses of SNA on the estimated latent dependency network among countries. One of the advantages of spatial econometric approach is that one can actually *estimate* the effective dependency structure among countries on the latent space using several observable networks. In my case, for example, geographical proximity, trade flows and regime similarity are observable based on the given data. However, after running the spatial model, we learn to what extent the three dependency networks actually contribute to states' regime influence. This is the *estimated latent (or implicit) network*. We have information about the tie strength in this implicit network in the form of adjacency matrix. This network can be plotted as a network graph, or its characteristics can be directly summarized in typical SNA indicators such as centrality scores and network degrees of various nodes. I believe this is a meaningful way to take advantage of the strengths both in spatial analysis (in which networks are estimated) and SNA (which provides a wide variety of summary index to describe given networks).

Appendix A.3.2 sketches a brief research note for the next iteration of analyses, extending the second simulation results. The simulation showed that systems with reinforcement mechanism, unlike systems without the mechanism, are sensitive to initial conditions. In the next step, I will slightly broaden the focus and ask whether this particular system is "history dependent", including "path dependent". Is democratization a path-dependent process? Scholars often casually state that institutions or political/economic regimes are path dependent (e.g., Pierson 2000). However, it is often unclear what they mean by path dependent, and there has not been rigorous statistical analyses for it. I employ Page (2006)'s rigorous mathematical definition of history dependence, which includes (i) initial condition sensitivity, (ii) *phat* dependence, and (iii) path dependence. The definition of the phenomena becomes stricter toward the latter. I suggest a use of a spatial-network hybrid statistical tool developed in Franzese, Hays and Kachi (2012) to test for the existence of history dependence in the context of political regime development.

The rest of the essay proceeds as follows. Section 4.2 reviews the earlier democratization literature, and Section 4.3 reviews existing studies on the geopolitics and the diffusion of democracy. Section 4.4 describes the main theoretical innovation of my arguments on regime-support networks and regime reinforcement. Section 4.5 explains how my theory is relevant to, and different from the conventional idea of diffusion, by introducing the concept of homophily and coevolution. Section 4.6 describes the final theoretical model for which

I test empirically in the later section. I will also present the main primary methodological challenge that arises from my theoretical model in Section 4.7. Section 4.8 introduces a new spatial econometric model that accommodates the methodological concerns listed in preceding sections. Section 4.9 describes concepts behind variables included in the model and how they are measured in data. Section 4.10 reports the estimation results. In Section 4.11, two sets of simulations demonstrate how small differences in initial conditions can make dramatic differences in the course of the system evolution over time. These results are at least as important as the estimation results reported in the preceding section, as they show clearly why my theory might be more realistic in explaining democracy levels of countries. Section 4.12 summarizes the main findings and the theoretical and methodological innovations of the study. Finally Appendix A.3.1 suggests some uses of the traditional social network analysis (SNA) approach in order to interpret and study further the latent regime-dependency network *estimated* in my spatial econometric model, and Appendix A.3.2 lays out the purpose and the research design of the next iteration of diffusion studies, which asks whether democratization is a path-dependent process in a mathematically and statistically rigorous manner.

4.2 Democracy Explained by Country Attributes

Many believe in the intrinsic and non-instrumental value of more democratic regimes as opposed to autocracies. The belief has been reinforced as we observed positive outcomes associated with democratization, such as higher economic development levels, lower levels of violence, and culture that values education, human rights, free media and so on. These elements have been considered as both potential causes and consequences of democratization. Academics have attempted to disentangle the causality among these social and political elements, and practitioners have been eager to assess whether these suspected causes of democracy have any significant effects.

Most of the earlier democratization studies focus on estimating the effects of domestic political and socio-economic factors on democracy. These studies theorized that either the level of democracy or the transition probability to democracy depends on various state-specific attributes. One of the most universally embraced links is that between economic development and democracy. This empirical question of whether economic development induces democratic transitions has been a main theme in this literature since Lipset (1959) first raised it half a century ago. Many scholars supported this idea of socio-economic development as a prerequisite of democratization. The seminal work by Huntington (1991) also advances a theory that the economic growth in the 70's and the resulting rise of the educated middle-class population led to the emergence of the "third wave" of democracies that started as late as the mid 1970's. Przeworski and Limongi (1997) and Przeworski et al. (2000) challenge this view, by empirically demonstrating that democratization occurs at random, and that a democracy sustains higher GDP levels ceteris paribus. Later some other scholars challenge this finding by providing the opposite result (Boix 2003; Boix and Stokes 2003). Epstein et al. (2006) also show that the effects of economic variables might be sensitive to how the dependent variable is coded. They show that the effects are not statistically significant with the dichotomous regime variable just as in Przeworski and Limongi (1997) and Przeworski et al. (2000), but they become statistically significant when political regimes are categorized into three levels. In my model, I consider several other country-specific attributes that are well-grounded in the literature. These include countries' own democracy scores from the previous time period, economic growth rates, commonwealth membership (or colonial history), urban population rates, fuel export rates (or resource richness), and religions.

4.3 Studies on the Diffusion of Democracy

Also in the 1990's, scholars of international relations and geography who study spatial patterns of political phenomena shifted attention to political outcomes such as regime type and war to test their hypotheses regarding the geographical clustering of these outcomes. Studies in geopolitics of democracies emerged independently of the studies of democratization with domestic causes, such as the ones mentioned above. Mechanisms aside, geographical studies of democracies, such as O'Loughlin et al. (1998); Starr (1991); O'Loughlin, Staeheli and Greenberg (2004), have revealed that similar political regimes tend to cluster both temporally and geographically, by careful tabulation, mapping and graphing of geographically subsetted data. This is when scholars started to ask what determines the seeming clustering of regimes, or what determines the local convergence toward the autocracy and democracy "clubs".

The theory of contagious political regimes emerged from these findings on the regime clustering. Scholars theorized that political regimes might be "contagious" across countries just as diseases and habits can travel across individuals who are well connected. The common motivation lying under these studies is that regime types cannot be accurately predicted if we only look at domestic characteristics of a single country or if we treat the level of democracy across countries as independent outcomes. Gleditsch and Ward (2006), for example, provide a model of regime transition—both transition to democracy and reversion to autocracy—including domestic and international country-specific attributes (logged GDP per capita, logged energy consumption, civil war, peace/war status) and the factors that somewhat captures the influence of other countries' regimes (global and local proportions of neighboring democracies). Kopstein and Reilly (2000) and Brinks and Coppedge (2006) also take similar approaches. These studies incorporate variables that measure both domestic and international determinants of democracy levels (or the probability of democratic transitions) in a single model.

However, all of these studies suffer from a couple of non-negligible limitations. First, substantively, the channel of regime contagion is assumed to be geographical proximity. Only the physical distance among countries is assumed to facilitate interstate influences of political regimes. Second, the way they implement "the influence of other states' regimes" in their empirical models is quite difference from what they refer to as "contagion" in their substantive theory. Contagion means a dynamic in which a country's democracy level *influences* others' democracy levels. The effects of other countries' regimes should not only depend on *how many* of other democracies a country is connected to, but also on (i) *exactly to which of others* one is connected and (ii) *how democratic each of the other countries is.* In other words, the concept of contagion should reflect the nature of *interdependence* across countries, but these earlier studies reduced the topological and relational information among countries down to a simple country-specific attribute that captures how democratic its surrounding is. Methodologically, this is exactly the same as other domestic determinants of democracies. In (social) network scientists' words, these are all "node attributes" and "the network tie" information is lost.

Beck, Gleditsch and Beardsley (2006) overcome these two shortcomings by applying a spatial econometric model to a cross-sectional data of democracy. First, as the title of their article "Space is More than Geography" implies, they warn us that geographical proximity is just one of many regime-dependence patterns that can facilitate contagion of political regimes. To prove this point, they incorporated a non-geographical concept of dependency network, economic dependency (measured by trade volumes), and showed that countries influence democracy levels in other countries more significantly when they are economically more dependent, as well as when they are geographically closer. Second, by using a spatial econometric approach, they were able to model explicitly a state's democracy level by other states' democracy levels, weighting the strength of others' influences by given connectivities across the countries (through trade and geography).[8] Goodliffe and Hawkins (2011) took a step further and considered three non-geographical dependency networks (trade, alliance, and international organization (IO) partners) through which democracy and autocracy might

[8] However it is also important to note that their empirical analysis was on a cross-sectional data for a single time point. This issue, cross-sectional vs. time-series cross-sectional will become relevant once I introduce yet another concept of contagion space, and the longitudinal data are introduced.

spread across countries.

4.4 A New Theory: The Homophilic Formation of Regime-Support Networks and Regime Reinforcement

Methodologies used in the existing studies attempting to test for the contagion of democracies (or autocracies) have a number of shortcomings, but these studies also offer insightful theories about the regime diffusion; *and yet*, in my view, even their definitions of dependency space is quite limited.[9]

In this essay, I re-examine mechanisms of regime diffusion by incorporating a different type of space, in addition to what existing studies implement, such as in Beck, Gleditsch and Beardsley (2006) and Goodliffe and Hawkins (2011). Just as in the existing diffusion studies, I first argue that countries that are more strongly dependent on each other influence each other's regimes more significantly: states' dependency structures affect their levels of democracy and autocracy. Based on the significance revealed in the existing studies, I decided to include geographical proximity and economic dependency (through trade) as spaces through which contagion occurs.

However, the main claim that makes my diffusion theory unique is that states can also strengthen other states' democracy/autocracy levels, either by an explicit economic support or an implicit diplomatic and political-ideological support. Because of the other states' support or approval, a country can potentially strengthen—or at least maintain—its democracy (or autocracy) level over time. This incremental influence between states can generate a

[9] There are also a couple of methodological issues about which my approach is more favorable compared to any of the aforementioned studies, but first, in this section, I elaborate on the substantive theoretical mechanism that motivates this paper.

process where ones that already have relatively similar regimes become even more similar in the long-run, all else equal—*regime reinforcement.* For a country that is already relatively democratic, a support from others might take a country's democracy score to a higher level, all else equal. For a country that is now relatively authoritarian, others' support could help her maintain the authoritarian reign or deepen it. In sum, my main theoretical argument consists of three closely connected components: (1) first, countries with more similar regimes form stronger dependency ties of political (diplomatic) and economic support, (2) second, through this dependency network based on support, the contagion of regimes occurs, and (3) as the consequence of the above two mechanisms (self-selection into support networks and the influence through the networks), countries' political regimes can be reinforced in the long run.

Political and diplomatic support among countries is rarely observable and the very concept of "diplomacy" contains a lot of nuances; however there are quantitative and qualitative evidences indicating that states tend to support each other diplomatically when their political regimes are similar. For example, Brinks and Coppedge (2006) point out the possibility that political leaders can use their neighbors' regimes as "good or bad examples" in favor of their own regimes, inducing the neighbor emulation effect in political regimes. Neumayer (2008), in his empirical study with 155 states, shows that the pairwise ideological similarity between two countries is an important determinant of the formation of diplomatic partnership network, along with geographical distance and the military power. In some rare cases, political leaders manifest regime alliances in the form of economic support *and* the information becomes publicly available to us. Here is an example of such an extreme observable case. In December 2011, the Brazilian Embassy in Santiago, Chile, disclosed as many as 266 telegrams sent to the Chilean governments during the period of 1973-1976. The telegrams revealed that the Brazilian military dictator Emílio Médici granted a fifty million US dollar worth of economic aid to the newly-emerged Pinochet government in 1973, soon after the coup de état in which Augusto Pinochet overthrew the President Allende (Pedigo 2011). The telegrams also reveal that, at one point, the Pinochet government explicitly asked the Médici government for its political support defending the Chilean coup in front of the international community.

These examples suggest that countries form diplomatic support/approval networks precisely when they have more similar—than dissimilar—regimes, selecting themselves into a regime-support network. Hence the strength of ties (or "edges", in the SNA language) of such regime-support networks should be defined by the degree of similarity in their democracy levels from the previous time period.

4.5 Comparison with the Conventional Concept of Diffusion

My theory of regime reinforcement is very much relevant to the existing concept of diffusion, and yet it is different. For this reason, my theory and the empirical test for it can be confusing. In this section, I elaborate on why the theory of regime reinforcement is similar to the existing diffusion literature but the dynamics generated in the two approaches are different.

First, my theory is relevant to the diffusion-of-democracy literature, because part of my theory is about regime influence across countries that occurs on the dependency network of regime support and approval. Just like countries' democracy/autocracy levels travel through dependency spaces defined by geographical proximity and the economic partnership, part of my claim is that countries' regimes influence each other through the dependency space defined by how similar states' regimes are. In this respect, my theory is also about an influence/diffusion/contagion mechanism.

In the following, I explain the differences between the conventional contagion process and my regime-reinforcement process in two ways—the short-run and long-run implications. These two together shape a dynamic called *network-behavior coevolution*, which is the fundamental dynamic that makes my theoretical model unique, and it raises a new set of methodological challenges.

4.5.1 The Short-Run Difference: Homophilic Selection

Why do I treat this particular type of dependency—regime-similarity network—differently from the ones used in the conventional diffusion studies, such as geographical connections and economic partners? The simplest answer to this question is that the construction of the two types of spaces is different. The former, the regime-similarity network, emerges from states' *homophilic self-selection*, while the latter does not.

Homophily is a phenomenon in which individual units form connections with similar others. McPherson, Smith-Lovin and Cook (2001) extensively review studies that demonstrate homophilic tie formation among individuals, which include situations where one becomes friends with others due to their similar behavioral habits. One of the archetypical examples would be where a smoker becomes friends with other smokers. In this example, a common behavior that connects these individuals is smoking, and the network formed in the process is a friendship network. The formation of regime-support network among countries is a homophilic dynamic in which states form invisible ties with others when their regimes are similar.

Scholars in SNA have been aware of the existence of homophily for quite a while (e.g., Lazarsfeld and Merton 1954; Coleman 1958), but only recently it has become a salient topic in SNA as a potential obstacle in making causal inferences on the association between networks and political behavior. For example, when we observe a cluster of smokers hanging out together, we can see the phenomenon in two ways. First, they might have become friends because they all smoke and see each other often outside the same building smoking. One might also guess that they were friends from the beginning for various reasons, and many of them picked up the habit of smoking cigarettes from one or two members of the group who were smokers from the beginning. In short, the question is: "did the friendship ("network") cause smoking ("individual behavior"), or did smoking ("behavior") trigger their friendship ("network")?" The former causal relationship is *contagion* and the latter relationship is *selection by homophily*.[10] Fowler et al. (2011) describe in detail the difficulty of distinguishing contagion from selection.[11]

Obviously, not all the dependency networks across states emerge from homophily. Networks among countries defined by geographical proximity are 100% given and fixed regardless of the countries' preferences. Likewise, numerous factors outside the state similarity characterize the trade network; for example, obviously it depend largely on the necessity and availability for certain goods and services, geographical distance which is exogenous and fixed, and some historical incidents that facilitate interstate transactions, such as the colonial history. None of these are purely correlated with the similarity in countries' political, economic and social choices.

4.5.2 The Long-Run Difference: From Homophily by Behavior-Type to Coevolution

To understand the true importance of my regime-reinforcement mechanism, however, it is necessary to consider the long-run consequence of incorporating the homophily-based dependency network (consistent with my theoretical argument), and compare it with the case without the homophily-based dependency network (consistent with the conventional

[10] This is self-selection, because individuals select themselves in the friendship network that consists of members who have a similar behavioral habit.

[11] Even though this causality issue has become salient and more scholars have started to pay attention to the problem only recently in SNA (as can be seen in the 2011 article by Fowlers et al.), one might want to note that Lazer (2001) already pointed out this issue in the SNA context back in 2001 and even suggested a solution based on a simpler and earlier spacial econometric approach. Also, distinguishing contagion and selection has been one of the major concerns among some spatial analysts for a long time, and the spatial-lag model developed in Hays, Kachi and Franzese (2010) is one of the first to suggest a simpler and tractable model that distinguished the two mechanisms.

concept of diffusion). To see the evidence of such a difference, we need to wait until the simulation analyses conducted in Section 4.11, but first in this section, I explain theoretically (i) what kind of difference one should expect to see from the two models, and (ii) why we should expect such a difference.

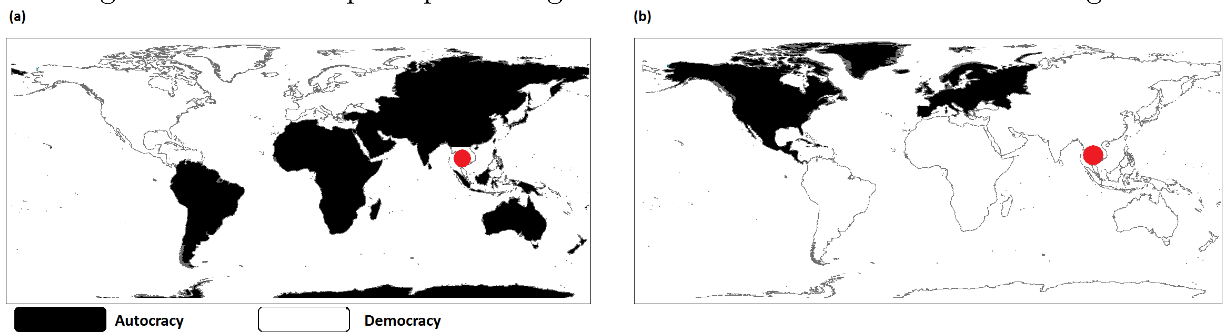
The one phrase that describes the difference is the "*initial condition sensitivity.*" The difference we should expect is that the world (or a "system", in a more abstract modeling language) is sensitive to a different surrounding conditions in an earlier time point, if we incorporate the homophily-based dependency across regimes. A world described without this type of dependency is not sensitive to initial conditions.

What does it mean to say that a system is or is not sensitive to initial conditions? To clarify, imagine two counterfactual worlds with very different regime distributions as depicted in Figure 4.4. The white part indicates democracy and black indicates autocracy in these counterfactual dichotomous worlds. Here is a scenario to consider. If a country indicated with a dot is born in the two different worlds—under the two different *initial conditions*, should we except that the country would experience the same regime development in the long run?[12] Should we expect that the country would eventually reach the same democracy level in the long run, starting from the two different initial conditions? As shocking as it might sound, by constructing a theoretical model *without* the homophily-based dependency, we would be assuming that the political regime of the "dot" would achieve exactly the same democracy level in the long run; in fact, we would be assuming that the political regimes of all the states included in the system eventually converges to the same (a single) level of democracy.

I would describe the puzzle of this initial condition sensitivity, or the puzzle of global convergence, as follows. First, in my view, it is very unrealistic to assume that a country's regime

[12]It is also essentially the same to ask "if there is *an exogenous democratic/autocratic shock to this country* under these two different initial conditions, should we except that the country would experience the same regime development in the long run?"

Figure 4.4: Two Maps Representing Different Distributions of Political Regimes



(even worse, all countries' regimes) reaches the same level of democracy, emerging under the two different sets of initial regime distributions. However, if one challenges me claiming he or she indeed believes that the two sets of initial conditions can eventually produce the same-looking world in terms of political regimes, I would still have to argue that one could confirm such a statement *only if* he or she allows for the possibility of democracies' diverging to different levels, starting their evolution under different initial conditions. Only then, if the term associated with the homophilic dependency turns out to be statistically insignificant, one could, for the first time, conclude that our world is quite simple and all the countries should eventually reach the same democracy levels, all else equal. This is why considering the homophily-based dependency is absolutely necessary for us, regardless of what one believes ex ante.

Now let us step back a little and examine why this difference exists between models with and without a homophily-based network. More precisely, not any kinds of homophily-based dependency can contribute to this difference: the key contributor is homophily that occurs *on the similarity of actors' behavioral type*. In the case of democracy, actors are counties, and behavioral types are political regimes. Note that this homophily occurs based on the similarity in the very behavior (regime choice) of countries' that is determined partially by the very dependency network of regime similarity.

In the longitudinal setting, the following is how regime influences occur among countries. First, countries with similar regimes form stronger dependency ties by homophily—homophily based on their behavioral type. This is self-selection. In the next time period, a state's regime influences that of other states' by the supporting or approving mechanisms. This is a diffusion mechanism. In the diffusion stage, countries with more similar regimes have stronger influences on each other's regimes, than countries with less similar regimes. After a round of regime diffusion, each country reaches a new democracy level. These new levels of democracy, in turn, define the updated regime-similarity network in the next time period. As a consequence, the latent connectivity across countries ("network") and their democracy level ("state behavior") co-evolve over time. This is the *network-behavior coevolution*. Network-behavior coevolution induces non-linearity in the system of 170 plus regimes. This systemic non-linearity is the source of initial condition sensitivity. To sum up, this is why a model with the particular type of homophily is sensitive to initial conditions, but a model without the homophilic dynamic is not.

4.6 A Unified Model of Democracy

I hope it is now clear that the regime-support network is significantly different from any other dependency networks included in the conventional diffusion studies. In this section, I will place my theoretical argument in a fuller context of modeling democracy levels, and present a unified model of democracies.

To repeat my research question, it is "what determines countries' democracy or autocracy levels?" The main dependent variable is the level of democracy. While maintaining the emphasis on the reinforcing mechanism of political regimes across countries, my predictive model for the level of democracy also incorporates important existing theories. I argue that there are three major components in a model determining the level of democracy.

First, as the earliest democratization literature suggests, the level of democracy largely depends on a country's own characteristics. For example, Beck, Gleditsch and Beardsley

(2006) include (only) the GDP level of each country in their empirical analysis. I decided to select several country-specific characteristics well-grounded in the exiting studies—mainly Przeworski et al. (2000). My model includes variables for countries' economic indicators, colonial history, urbanization indicator, resource richness, and religious profile.

The second mechanism is the diffusion process, in which a country can influence other's regimes through pre-defined dependency networks. The interstate influence is stronger when the countries are more dependent on each other, or the tie between the countries is stronger. This mechanism of diffusion reflects the concept of spatial autocorrelation summarized succinctly in Waldo Tobler's First Law of Geography; "everything is related to everything else, but near things are more related than distant things."

Some criticize that diffusion studies lack substantive theories. I have never fully agreed on this criticism, but I do notice that some diffusion studies focus so much on explaining the mechanical working of contagion that they fail to provide the explanation of what "influence" can mean in each substantive context. I need to warn as a spatial analyst, however, that just as it is true to many empirical analyses, there is no mechanical statistical "test" that can tell us exactly "what the nature of influence" is in regime contagion. Only careful substantive-theoretical studies can tell us why diffusion occurs and what the nature of influence might be in the given context. Spatial econometric models, for example, can then tell us whether these theoretically suggested dependency networks matter at all in explaining observed clusters of political outcomes. Here are some plausible mechanisms behind the contagion of democracies and autocracies that I believe are at work.

First I list four possible distinct natures of influence among individuals or countries in political science, using the categorization in Simmons, Dobbin and Garrett (2006). They are coercion, competition, learning and emulation. A country could be forced to take or not take a certain political action under some external pressure (coercion). A political actor might take a certain action whenever their competitors take certain actions (competition).

An actor can learn from other actors' experience about the effectiveness of newly adopted behavior (learning). Or actors' nature of attempting to produce better outcomes can make an actor mimic others' actions (emulation).[13]

I argue that learning is a predominant mechanism behind regime contagion. For example, if citizens in these countries learn about the odds of successful revolution against the authoritarian government by observing the masses' behavior of other countries, revolution might "spread" around the world, leading the movement to democratization of the country. Studies show that citizens do learn rationally from successful revolutionary uprisings in other nations (e.g., Weyland 2009).

Therefore, actual dependency networks included in empirical models (i.e., observable dependencies) should be ones that are suspected to facilitate this type of learning. It is not hard to imagine that geographical proximity is a promising candidate. For example, people who live under an autocratic regime most likely tend to gauge the odds of success from an uprising in a neighboring country to ones in distant countries, everything else equal. The economic dependency is also often a pathway for any political information. Both at the elites' and citizens' levels, countries know each other better when there are stronger economic ties. This is the reason my model, just as most other diffusion studies, includes the economic dependency as a candidate of a pathway for regime influence.

Finally, the third component of my theoretical model reflects the regime reinforcement theory, where countries influence each other's regimes when the two regimes are more similar than dissimilar. As I fully described the argument in the previous sections, the reason for the regime influence is countries' support or approval of other similar regimes. Under Simmons, Dobbin and Garrett (2006)'s categorization, the regime support and approval mechanism most likely belongs to emulation.

[13] An earlier work by Most and Starr (1990) also presents a typology of spatial associations in the outcome variable and suggested five distinct mechanisms: societal, reinforcement, extra-societal, imposition, and selective. However, some of them are not precisely diffusion.

Mohrenberg (N.d.) explores a similar theoretical idea in his working paper. What the author refers to as "diplomatic relations" is essentially dependency networks defined by countries' regime similarity. He asks whether contagion of democracies occurs through this diplomacy network. The idea in his study is also similar to my strategy, in that the author attempts to model diffusion as a truly relational mechanism unlike other existing "diffusion" studies. Unfortunately, the author develops a theory of self selection, but does not provide empirical analyses. However, the author suggests an empirical strategy at the end of the paper. Instead of taking the spatial-econometric approach, this study suggests the use of *SIENA* (developed by Snijders and colleagues), an SNA-based statistical tool that combines contagion and selection (Snijders 1997, 2001, 2005; Snijders, Steglich and Schweinberger 2007).

Among the social network analysis approaches, *SIENA* is the most advanced and perhaps the only methodological solution for modeling coevolution, and in combining contagion and selection. However, the construction of their coevolution model is quite hard to understand and as far as I understand, there are quite a few limitations in using their estimation strategy. What is rather inconvenient for analysts is that the developers state in quite a few places that there are aspects of their estimator performance that are unknown or problematic.[14]

Before I move on to list methodological challenges, I would like to note, however, that the homophlic tie formation between similar regimes is essentially the same as what that "one flap of a butterfly's wings" does. It is a small incremental event that can cause a significant differences to the course of the system evolution later.For this reason, I do not expect that regime-reinforcement among states would have a dramatic effect on the level of democracy within a single time period, after controlling for the effects of state-level characteristics (the GDP level etc.) and the effects of others' regimes through conventional dependency paths. After all, the support of other regimes is a subtle concept. However, this seemingly small dynamic of states' self-selection into the reinforcement network is just like a flap

[14]This issue is discussed further in Franzese, Hays and Kachi (2012).

of a butterfly's wings. As long as the reinforcement dynamic exists, or as long as it is "statistically significant" in econometricians' language, the system based on my theory of democratization demonstrates a significantly different story in the long run from what we already know based on existing studies. The difference is mainly its sensitivity to the initial world-wide distribution of democracies. I will come back to this point when I discuss my simulation studies.

4.7 Methodological Challenges

Do data support my theoretical arguments? To conduct an empirical analysis for my theory of democracies, there are mainly two methodological difficulties to be conquered. One corresponds with a typical shortcoming of existing diffusion studies, and the other emerges from incorporating the new theory of regime reinforcement. The use of the multiparameteric spatiotemporal autoregressive model with coevolution (MSTARC) provides a solution for the two problems simultaneously.

4.7.1 Theories of Relations, Inferences of Attributes

First, what claims to be a study of diffusion often develops substantive theories that involve *influence* mechanisms. In my study, for example, a country's democracy level influences others' democracy levels through geographical, economic, and regime-support dependency networks. In other words, the degree to which country A and B are dependent on each other determines the degree of regime influence between the two countries. Imagine a dependency network depicted in the figure below (Figure 4.5). The diffusion theories make claims about the effects of the relational information contained in ties (or "edges" in the SNA language) on the node/unit behavior or attribute. In the figure, the size of edges is the strength of dependence between countries, the color of nodes on the right-hand side represents node attributes (e.g. as the GDP level of each country), and the node shape on the left-hand side indicates the explained individual behavior/attributes (e.g. different levels of democracies or different regimes).

Figure 4.5: The Effects of Tie/Edge/Relatinal Characteristics

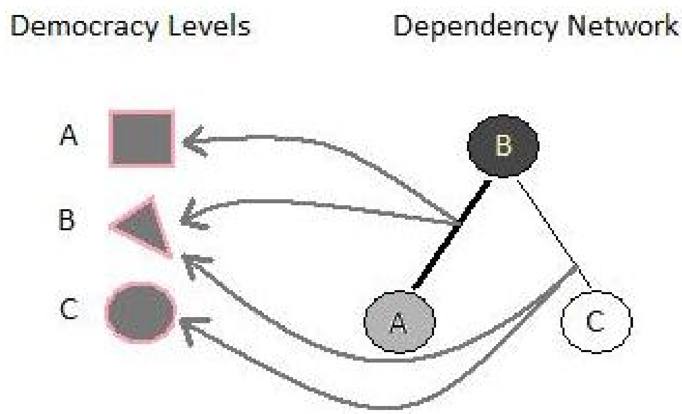
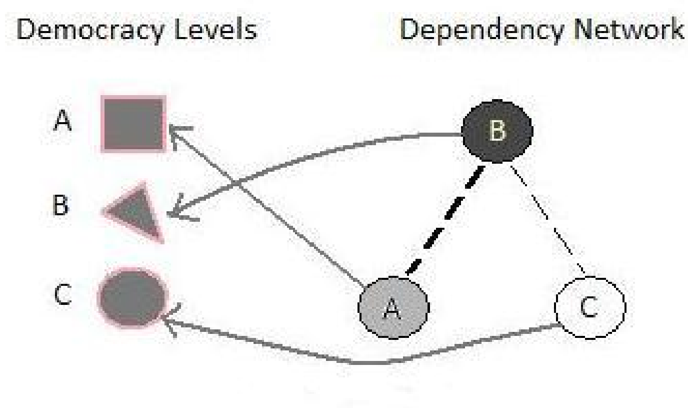


Figure 4.6: The Effects of Node Characteristics



However, most of the empirical models in existing diffusion studies end up reducing the tie/edge/relational information down to the country-specific attributes (e.g., Gleditsch and Ward 2006; Brinks and Coppedge 2006). For example, a common strategy is to create a country-specific score that measures "what percentage of the country's geographical neighbors is democracy." These studies start with a relational theory (through geographical proximity, for example) but end up implementing the "diffusion mechanism" as a country/node-specific attribute, just as other covariates such as economic indicators and the religion profile of the country. Figure 4.6 depicts this type of empirical model. This empirical model does not contain the relational information among counties anymore. (Note that the edges drawn in this figure do not paly any roles.)

More technically speaking, the neglect of the true dependence structure is becomes a source of estimation biases—simultaneity biases. If such dependence truly exists, then there is a feedback loop across countries' democracy levels. Suppose there was a positive shock in country A's economic variable, and suppose it increase A's democracy score. If there is a true, and say positive, influential path between A and B (as in the picture), then B's democracy score increases. Since B is also connected with country C, C's score also improves. The flow of regime feedback does not necessarily stop at this point. The influence will goes back and forth from A to B to C to B to A to... until the whole world eventually reaches the long-run convergent point, the steady state.

4.7.2 Distinguishing Selection from Contagion

The second methodological challenge stems from incorporating the regime-support network, which is a self-selective network by countries with similar regimes; more precisely the selection occurs by homophily. As explained earlier, homophily that occurs by the similarity in the unit behavior (i.e., countries' democracy levels in this study) generates a distinct feedback loop across time and space. In short, the network induces the regime influence (reinforcement) affecting each country's democracy level, but the regime levels also define the support network in the next time period—network-behavior coevolution. One way to see this challenge is that what we thought was the contagion effect before might partly be a mere consequence of self-selection. If self-selection truly exists, then we would overestimate the contagion mechanism. How could we distinguish selection from pure contagion mechanisms that occur though geographical and economic dependency networks? As Fowler et al.

(2011) mention, causality between networks and individual behavior has been recognized as an important and yet very difficult problems to solve statistically.

The new spatial econometric model, MSTARC, that I introduce in the next section overcome both problems at the same time.

4.8 The Empirical Strategy

The empirical analysis is to assess how states' democracy levels influence those of other states' after controlling for the effects of country-specific common conditions. Furthermore the model needs to be able to capture the two kinds of diffusion effects. To answer these questions, I estimate a multiparametric spatio-temporal autoregressive (M-STAR) model with co-evolutionary dynamics. This is an extension of existing M-STAR models, which contain multiple spatial lags but not the term that captures the co-evolutionary dynamic. Hays, Kachi and Franzese (2010) provides technical descriptions of the empirical model in detail.

The key properties of the M-STAR-plus-co-evolution model are the following. First, the model in matrix notation is

$$\mathbf{y} = \Big[\sum_{r=1}^{R} \rho_r \mathbf{W}_r\Big]\mathbf{y} + \phi\mathbf{M}\mathbf{y} + \gamma\mathbf{L}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \varepsilon, \tag{4.1}$$

where $\mathbf{y}$, continuous democracy scores, is an $NT \times 1$ column vector of cross sections stacked by periods. Throughout the paper, $i = \{1, \cdots, N\}$ denote countries and $t = \{1, \cdots, T\}$ denote time periods included in the data. Note that the dataset does not have to have the balanced-panel structure; i.e., the number of countries included in a given time period, N_t, could vary over time due to data availability and deaths and births of countries. In total, $\sum_t N_t$ observational units are included in the entire dataset. For simplicity (in terms of writing), however, I will describe the model as if the dataset had a balanced-panel structure, denoting the total number of observations by NT. This does not change the properties of this estimator.

The notation $\boldsymbol{\beta}$ on the right-hand side captures the effects of various country-specific covariates included in $\mathbf{X}$. If there are K covariates in total, then β is a $K \times 1$ column vector and $\mathbf{X}$ is an $NT \times K$ matrix. In my specification, $\mathbf{X}$ contains *logged GDP per capita* and the *constant* term. If any of these explanatory variables happen to be spatially clustered, then the term $\mathbf{X}\boldsymbol{\beta}$ controls for the seeming spatial correlations observed in the outcomes variable, $\mathbf{y}$.

Similarly, ϕ represents the effect of *one-year lag of a country's own democracy score*. Since this first-order temporal lag is a country-specific exogenous (pre-determined) variable, it can also be treated as one of the $\mathbf{X}$ variable and can be included in the $\mathbf{X}\boldsymbol{\beta}$ term. The time-lag matrix $\mathbf{M}$ is an $NT \times NT$ matrix that conveniently maps y_{it} onto i's own past value $y_{i,t-1}$. This notation allows us to use the vector $\mathbf{y}$ instead of $\mathbf{y}_{t-1}$ and this will provide us with a more intuitive expression of the reduced form equation later.

The first term on the right-hand side, $\left[\sum_{r=1}^{R} \rho_r \mathbf{W}_r\right]\mathbf{y}$ captures the exogenous diffusion. The indicator $r = \{1, \cdots, R\}$ denotes different kinds of connectivity through which a country's regime influence others'. Suppose we have a theory that the economic interdependence measured by trade volumes partially determines to what extent a country is influenced by the others' political regimes. Each cell, w_{ij}, of the observable and exogenous weights matrix $\mathbf{W}_{trade}$ contains the pair-wise trade volume between country i and j. Since the trade volume is defined as the sum of import and export, the $\mathbf{W}_{trade}$ matrix is symmetric. For example, suppose there are only three countries, A, B, and C, and the trade volumes among these countries are as in Table 4.2. Table 4.2 implies that the extent to which B and C's regimes influence A's regime is 3/10 and 7/10 respectively.

Table 4.2: Imaginary Trade Volumes among Countries A, B and C

Country	A	B	C	Row total
A	0	3	7	10
B	3	0	5	8
C	7	5	0	12

Then the weights matrix that corresponds with the chart in Table 4.2 is as in equation (4.2). Note that conventionally we row-standardize the values in weights matrices. As can be seen in equation (4.2), ρ_r is the estimated coefficient for the effects of others' regimes through this particular type of connectivity, r. In my specification, there are two kinds of exogenous-diffusion channels: one is trade volume and the other is border-sharing. In other words, in the model I estimate, $R = 2$. Note that, even though $\mathbf{W}$'s are exogenously given and the coefficient that we estimate is only the ρ, the overall effect of other countries' regimes (through one kind of connectivity) should be thought as the product $\rho_r \mathbf{W}_r$.

$$\rho_{trade,t} \mathbf{W}_{trade,t} \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \rho_{trade,t} \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.375 & 0 & 0.625 \\ 0.583 & 0.417 & 0 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix}. \tag{4.2}$$

The term $\gamma \mathbf{Ly}$ captures the effects of the endogenous diffusion. The matrix $\mathbf{L}$ is an $NT \times NT$ "regime-distance" matrix with $|y_{i,t-1} - y_{j,t-1}|$ in cells (it, jt). The $\mathbf{L}$ matrix plays the role of a weights matrix just as a $\mathbf{W}_r$ does in the exogenous-diffusion term. The difference is that each element of $\mathbf{L}$ is defined as a distance between the political regimes of each pair of countries. Adding this term, $\gamma \mathbf{Ly}$, therefore reflects a substantive proposition that countries with more dissimilar political regimes affect each other's political regimes more if $\gamma > 0$, and less if $\gamma < 0$.

The reduced form equation follows from equation (4.1);

$$
\begin{aligned}
\mathbf{y} &= \Big(\mathbf{I} - \sum_{r=1}^{R} \rho_r \mathbf{W}_r - \phi \mathbf{M} - \gamma \mathbf{L}\Big)^{-1} \Big(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\Big) \\
&= \mathbf{A}^{-1}\Big(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\Big),
\end{aligned}
\tag{4.3}
$$

where the matrix $\mathbf{A}$ is defined by $\mathbf{A} = \Big(\mathbf{I} - \sum_{r=1}^{R} \rho_r \mathbf{W}_r - \phi \mathbf{M} - \gamma \mathbf{L}\Big)$.

Finally, the likelihood function is

$$
L(\sigma, \boldsymbol{\rho}, \phi, \gamma, \boldsymbol{\beta} | \mathbf{X}, \mathbf{y}) = |\det \mathbf{A}| \Big(\frac{1}{\sigma^2 2\pi}\Big)^{\frac{NT}{2}} \exp\Big(- \frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\Big), \tag{4.4}
$$

assuming that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT}), i.i.d.$

4.8.1 The Long-Run Implications of the Co-Evolutionary Dynamic

The methodological complication and the interesting dynamic caused by the co-evolution between $\mathbf{y}$ and $\mathbf{L}$, which never existed in the traditional spatial models, becomes evident when we rewrite the model for a cross-section given a time-period t. First, the structural-form equation for a cross section can be written as;

$$
\mathbf{y}_t = \sum_{r=1}^{R} \rho_r \mathbf{W}_{rt} \mathbf{y}_t + \phi \mathbf{y}_{t-1} + \gamma \Big[abs\Big(\boldsymbol{\Pi}(\mathbf{y}_{t-1} \otimes \mathbf{I}_{N_t})\Big)\Big] \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \tag{4.5}
$$

and the reduced-form cross-sectional equation directly follows from (4.5);

$$
\mathbf{y}_t = \Big(\mathbf{I}_{(N)} - \sum_{r=1}^{R} \rho_r \mathbf{W}_{r,t} - \gamma \Big[abs\Big\{\boldsymbol{\Pi}(\mathbf{y}_{t-1} \otimes \mathbf{I}_{(N)})\Big\}\Big]\Big)^{-1} \Big(\phi \mathbf{y}_{t-1} \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\Big), \tag{4.6}
$$

where $\mathbf{y}_t$, $\mathbf{W}_{r,t}$ and $\mathbf{X}$ are $N \times 1$, $N \times N$ and $N \times K$ matrices. The matrix $\mathbf{\Pi}$ is $N \times N^2$ and it is produced by horizontally concatenating N separate $N \times N$ block matrices. The ith $N \times N$ matrix in $\mathbf{\Pi}$ has -1's on its diagonal and 1's for each element of the ith column except for the element (i, i), which is 0 as are all other unspecified elements in $\mathbf{\Pi}$. For example, if $N = 3$,

$$\mathbf{\Pi} = \left(\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & 1 & -1 & 0 & 0 & 0 \end{array}\right). \tag{4.7}$$

As can be seen in equation (4.6), the presence of the regime-distance connectivity, $\mathbf{L}$, renders the N-equation system nonlinear in the endogenous variable, $\mathbf{y}$. This significantly complicates calculations of the predicted effects. In fact, there is no analytical solution for the steady-state regime levels anymore, and the long-run responses (democracy levels) to changes in the covariates $\mathbf{X}$ must be calculated recursively. This is why the predicted long-run democracy levels are path dependent and the level(s) to which the system converges varies across different starting values of $\mathbf{y}$.

It is useful to compare this model with the one for a simple M-STAR model that does not contain the endogenous-diffusion term. The structural form of (a cross section of) the simple M-STAR can be written as follows. Note that there is no term that represents the endogenous diffusion.

$$\mathbf{y}_t = \sum_{r=1}^{R} \rho_r \mathbf{W}_{rt} \mathbf{y}_t + \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t. \tag{4.8}$$

Unlike equation (4.5), one can analytically compute the steady-state (long-run) outcomes by equating $\mathbf{y}_{t-1}$ to $\mathbf{y}_t$ and fixing the exogenous right-hand-side variables to their counterfactual permanent post-shock levels;

$$\mathbf{y}_t = (\mathbf{I} - \rho \mathbf{W} - \phi \mathbf{I})^{-1} (\mathbf{X}_t \boldsymbol{\beta} + \varepsilon_t). \tag{4.9}$$

4.9 Variables and Data

4.9.1 Dependent Variable: Level of Democracy

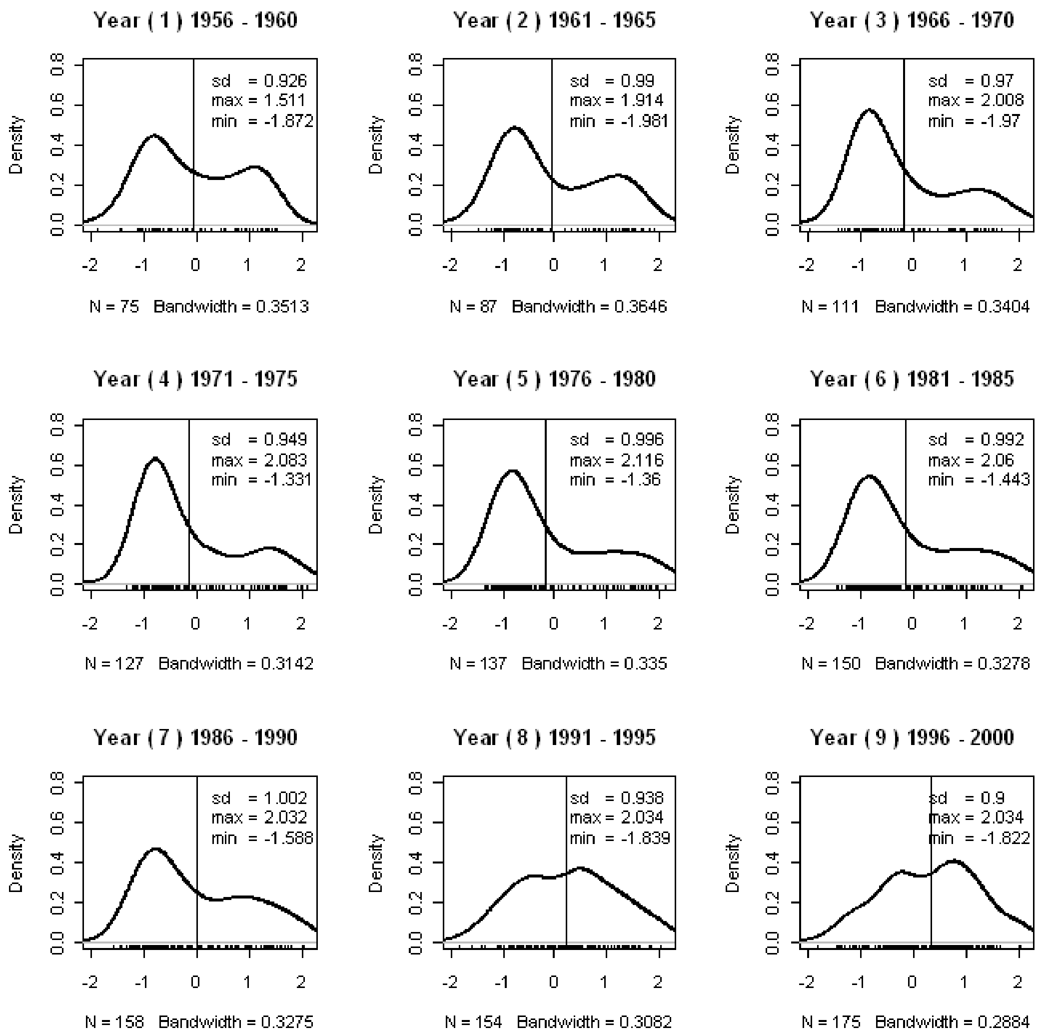
The dependent variable is the level of democracy and the data are from the Unified Democracy Scores (UDS) constructed in Pemstein, Meserve and Melton (2010). A Bayesian latent variable approach, originally introduced in Treier and Jackman (2008), allows them to estimate the latent levels of democracy, integrating ten different existing democracy scores[15]. The UDS data provide a continuous measure of democracy levels in 67 to 191 countries, depending on years, from 1946 through 2000. For the empirical analyses here, I use UDS for the 1950-2000 time-period, and take the 5-year average to create a 10-time-block time-series-cross-sectional regime data. Taking the 5-year average has a couple of advantages. One is simply to reduce the dimension of the data set. With hundreds of observational units within each year, the matrix dimension, *(the number of observations across time and units)*×*(the number of observations across time and units)* can easily become very large.[16] More importantly, by taking the average, I could alleviate possibly erroneous fluctuations of democracy scores, which is a major concern often raised by the believers of dichotomous or trichotomous democracy scores as an argument against continuous scores. Figure 4.7 demonstrates summary statistics of the democracy score for each of the 9 time-periods.[17] The scores range approximately from -2 to 2, and the means (indicated by vertical lines) are

[15]The ten measures are Arat (2003), Bowman, Lehoucq and Mahoney (2005), Bollen (2001), *Freedom in the World 2007* (2007), Hadenius (1992), Przeworski et al. (2000), Marshall and 2006 (N.d.), Coppedge and W.H. (1991), Gasiorowski (1996), Vanhanen (2003).

[16]For example, suppose the dataset includes 50 years and each year includes 100 countries. The dimension of weights matrices would become 5000-by-5000 and it incurs a substantial computational burden. It did not appear to me as a good strategy at this preliminary stage of the project to use such a large dataset.

[17]The ten time periods that I created by taking the five-year averages are; (0) 1951-1955, (1) 1956-1960, (2) 1961-1965, (3) 1966-1970, (4) 1971-1975, (5) 1976-1980, (6) 1981-1985, (7) 1986-1990, (8) 1991-1995, (9) 1996-2000. As I explain later, one of the explanatory variables is a one-time-period lag of the dependent variable. Consequently time-period (0) drops out of the dependent variable, leaving nine time periods, (1)-(9), in the final dataset.

Figure 4.7: Summary of the Democracy Scores (Dependent Variable) by Time Period



Note: Data source: Pemstein, Meserve and Melton (2008) roughly around 0 in all time-periods.[18] To get a sense of how countries are ranked in a given year, I plotted the democracy score of each country for the last time period, 1996-2000 (or "year 9"), in Figure 4.8.

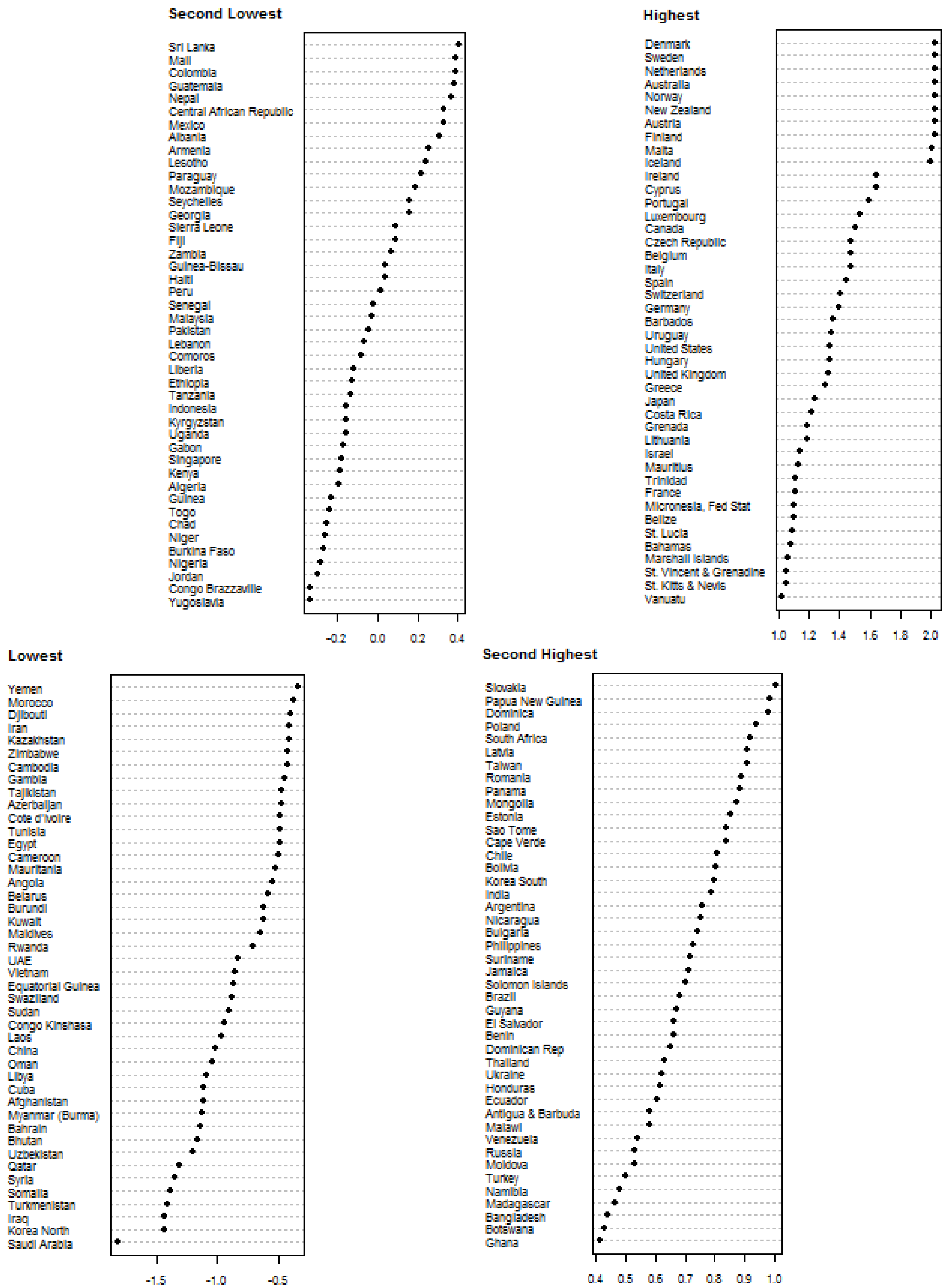
4.9.2 Explanatory and Control Variables

Each of the two main model specifications consists of three parts. First, following Lipset's social and economic requisites theory, the models contain country-specific social and economic attributes. A number of earlier democratization studies argue that the economic development level is positively associated with countries' democracy scores(Huntington 1991; Boix 2003; Boix and Stokes 2003). To allow the possible curve-linear relationship between the economic development and the democracy levels, I include *real GDP per capita* and the squared term, *(real GDP per capita)*^{2}.[19] The data are taken from Gleditsch (2002) and it measures the real GDP per capita in thousands of constant U.S. dollars with the base year being 1996. Another economic variable is *growth rate*. Scholars have emphasized the importance of growth in explaining democratization and consolidation of democracies, but have found weak or no empirical evidence of economic growth. It might be due to multiple competing forces that growth can generate. As Boix (2003) theorizes, "[t]he possibility that low taxes may spur faster economic growth may entice the poor to commit to moderate levels of redistribution" and that "'should, in turn, reduce the wealthy's opposition to universal suffrage and hence facilitate the introduction of democracy." With this mechanism, economic growth should be positively correlated with the level of democracy. As the author points out, this mechanism works only when the country has some political institution that ensures that the poor abide by their commitment to maintain the taxes low once democratization occurs.

[18]Unlike commonly-used democracy scores, such as Polity IV, the UDS are not restricted to a certain range. Each country's democracy score, estimated as its posterior mean, happens to have fallen between about -2 and 2.

[19]Later I tried estimating the models with the logged real GDP/capita variable instead of GDP/capita and (GDP/capita)^2. Qualitatively the results remained unchanged.

Figure 4.8: Democracy Scores of the 175 Countries included in the 1996-2000 Time Block



Note: Data source: Pemstein, Meserve and Melton (2008)

However, another possibility is that growth might reduce the level of economic grievances among the repressed. With this mechanism, economic growth should reduce the force to mobilize population and should be negatively correlated with the occurrence of democratization. I should note that this demobilization force can be realized only in the authoritarian regimes that have some (imperfect) political institution that ensures redistribution not only among the elites but also among the poor as well. The variable, *growth rate*, is computed as the log difference of the real GDP levels. The last economic variable is *fuel export* that measures the percentage of a country's fuel export of all the merchandise exports. The data are taken from the World Development Indicator (the World Bank). When a country democratizes, it usually means that the government enfranchises the poor/repressed. This moves the median voter position from somewhere in the wealthy group to a point in the then repressed group, resulting in higher taxes for the rich than in the former authoritarian regime (Boix 2003). While owners of relatively mobile business, such as manufacturers, could move their production sites outside the country if higher taxes are implemented, fuel (or natural resources in general) is a highly location-specific asset and it would be difficult for the rich to "move" the production sites when the tax rates increase (Boix 2003). Therefore in the fuel-rich countries, the rich have an incentive to block democratization movements and we should expect the negative effect of the *fuel export* rate on democracy scores.

The *commonwealth* variable is a dummy indicating the commonwealth membership. This variable is to capture the possible legacy of British colonization.[20] Due to the imposed British political institutions during the colonial era, these countries might have some political attributes in common with Britain that might lead to stable democratic regimes like the British one, regardless of their own political history before or after the colonization. Another socio-political variable is *urban population rate*. I expect that the urban population rate would be positively correlated with democracy for a couple of reasons. First, the

[20]Later I should change this to a variable that indicates only former British colonies and not the commonwealth membership.

urban population is more likely exposed to foreign culture, among which can be democratic countries. Second, higher degrees of urbanization imply higher population concentration. This could facilitate, in conjunction with the exposure to more democratic foreign culture, solving the collective actions that are oftentimes a key to overturn the dictatorial incumbents either by voting for the democratic opposition or revolting against the government. A concern might be that the urban population rate tends to increase as a country experiences economic development and it might be difficult to distinguish the effects of the urban population rate and other economic-status variables. Following Przeworski et al. (2000), three religion variables– *catholic*, *muslim* and *protestant*– are also included. Each measures the time-invariant percentage of the population that belongs to the religious group. The data are mainly from Przeworski et al. (2000) and for the countries that are not included in Przeworski et al. (2000), the current issue of the CIA World Factbook are used. As Przeworski et al. (2000) mention by citing Lipset (1959), Protestantism's emphasis on individualism and self-reliance is said to nurture democratic values while Catholicism "was antithetical to democracy in pre-World War II Europe and Latin America." Therefore we should expect to observe the positive association between *protestant* and democracy, but negative between *catholic* and democracy.

Lastly, the *temporal lag* variable is a one-year lag of a country's own democracy score. It controls for the history of the country's regime. If a country has a fairly democratic regime in one time period, then it is most likely that the democracy score of the country in the following time period is not too far away from the past score.[21]

It is important to control for these country-specific attributes not only because of each substantive rationale that I mentioned above. It is also because controlling for some of

[21] Overall the set of these country-specific variables is very similar to the one Przeworski et al. (2000) have in one of their fuller models. The differences are that my specifications do not include variables that capture social fragmentation, such as ethnic and religious fragmentation, and also that I control for the effects of fuel export when they do not. Another important source of social divide that I do not have in my current specifications is income inequality, which is a key variable in Boix (2003). I plan to include these variable as soon as I collect them.

these socio-economic variables is essential to extract the "true diffusion" or "true" spatial-interdependence of the democracy level, which is the central theme of this project (Franzese and Hays 2006, 2007, 2008). What does this mean? Supposed that we believe political regimes diffuse among countries through geographical connectivity; i.e., a country's democracy level is affected more by that of geographically closer others than that of distant others. At the same time, suppose it is the case that the similar levels of GDP/capita are geographically clustered. Finally suppose that we observe geographical clustering also in the dependent variable– the democracy score– as well. Now for the geographical clustering of democracy scores, two completely different explanations can co-exist. One is that political regimes spread among neighbors based on the geographical proximity of the countries. When there is an increase in the democracy score of a certain country, the democratic shock affects closer countries more than distant countries. This is one way we can observe geographical clustering in the regime score, and we consider this as a "true" diffusion or interdependent mechanism. However, the same clustering in the dependent variable can also be observed when each country's political regime responds similarly to a similar economic development level– one of the country-specific variables. In this mechanism, countries are responding to their own levels of the economic variable independently, and there is no diffusion of regime type. In this case, what we observe is, in fact, mere spatial clustering/association of the dependent variable, but it is not diffusion or interdependence. In the real world the two mechanisms can co-exist, and therefore it is important to include variables that capture both mechanisms in order to distinguish the two.

The second category of the explanatory variables is exogenous connectivity. As exogenous regime-diffusion channels, the model includes *trade volume* and *border*. Each of these weights matrices is multiplied by the outcome variable, democracy score. The overall product of the weights and the outcome variable represents other countries' democracy levels weighted by the strength of ties between each pair of countries. What is estimated in the regression models is the coefficient parameter attached to the whole product. This parameter captures how much the particular kind of connectivity matters in the context of democracy diffusion. In other words, for a certain pair of countries A and B, even if B's democracy score is very high and the tie between A and B is very strong, we should conclude that the diffusion through this particular connectivity does not exist, if the coefficient parameter attached to this term is estimated to be statistically not significant. The connectivity matrix *trade volume* is a proxy for the economic interdependence among the countries. The rationale to include the trade weights is that countries' political conditions tend to become more similar when they are economically interdependent. Each entry of the trade weights measures the sum of the import and the export (in millions of current-year U.S. dollars) for each pair of two countries and the data are taken from Gleditsch (2002). In some existing studies, trade volumes are divided by the effect-receiving countries' total GDP levels. The logic behind this practice is that the influence through trade volumes depends on how significant the volume is compared to the overall economic size of the effect-receiving country. A disadvantage of this practice is that it becomes unclear as to exactly which affects the outcome quantity, an increase/decrease in trade volumes or a decrease/increase in GDP. For this reason, the trade measure in my empirical analyses is a simple sum of import and export. The other connectivity matrix *border* is different from the trade matrix in that all the entries of the matrix are binary, 0 or 1. In this preliminary empirical study, the geographical contiguity is defined strictly as countries that share inland borders. The data are taken from the "Direct Contiguity Data, 1816-2006 (Version 3.1. Online: `http://correlatesofwar.org`)" in the Correlates of War Project (Stinnett et al. 2002). The most strict definition of contiguity ("type 1" in the COW dataset) is used.[22] Both *trade volume* and *border* matrices are row-standardized.

The third and the most important category of the explanatory variables is the influence of others' regimes through the endogenous tie strength. The connectivity matrix for the

[22]By this definition, for example, Korea and Japan, which are separated by the sea are not "neighbors". This definition seems too strict for the purpose of this study. In the next iteration, I am going to use a different contiguity definition, or the distance between countries as Beck, Gleditsch and Beardsley (2006) do.

endogenous diffusion consists of the distances of pair-wise regimes. Each cell (it, jt) contains the value $|y_{i,t-1} - y_{j,t-1}|$, where $y_{i,t-1}$ are taken from the UDS. The endogenous connectivity matrix is also row-standardized.

All of the regression models include time and region dummies. The nine time dumies are for the periods of (1) 1956-1960, (2) 1961-1965, (3) 1966-1970, (4) 1971-1975, (5) 1976-1980, (6) 1981-1985, (7) 1986-1990, (8) 1991-1995, (9) 1996-2000, and the dummy variable for the last time period, (9), s dropped from the regression equations. The eight region dummies are for (1) Africa, (2) North America, (3) Central and South America, (4) Asia, (5) Middle East, (6) Western Europe, (7) Eastern and Central Europe and (8) Oceania. The eighth regional dummy for Oceania is dropped from the regression equation.

4.10 Estimation Results

What country-specific political and economic conditions lead to more democratic regimes? Are there diffusion effects in political regimes? Are countries' political regimes more likely affected by countries with already similar regimes, or dissimilar regimes? Table 4.3 reports the results of the maximum likelihood (ML) estimation of M-STAR models with the co-evolutionary dynamic.

Recall the regression equation;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \phi\mathbf{M}\mathbf{y} + \Big[\sum_{r=1}^{R} \rho_r \mathbf{W}_r\Big]\mathbf{y} + \gamma\mathbf{L}\mathbf{y} + \boldsymbol{\varepsilon}. \tag{4.10}$$

With the actual variables and the connectivity matrices that are included in the model, the following model is fitted.

$$\begin{aligned}\mathbf{y} =&\Big\{\beta_{cons}\mathbf{1}_{cons} + \mathbf{x}_{c.specific}\boldsymbol{\beta}_{c.specific} + \phi\mathbf{x}_{lagged\ y} + \mathbf{X}_{time\ dumms}\boldsymbol{\beta}_{time\ dumms} \\ &+ \mathbf{X}_{region\ dumms}\boldsymbol{\beta}_{region\ dumms}\Big\} \\ &+ \Big\{\Big[\rho_{border}\mathbf{W}_{border} + \rho_{trade}\mathbf{W}_{trade}\Big]\mathbf{y}\Big\} + \Big\{\gamma\mathbf{L}_{dissimilar}\mathbf{y}\Big\} + \boldsymbol{\varepsilon},\end{aligned} \tag{4.11}$$

where $\mathbf{y}$ is the continuous democracy scores from the UDS. The terms in the first curly brackets are for the country-specific variables, including time and region dummies. The matrix $\mathbf{X}_{c.specific}$ contains all the non-diffusion determinants of democracy, such as *real GDP/capita*, *(real GDP/capita)*^{2}, *growth rate*, *commonwealth*, *urban pop rate*, *fuel export*, *catholic*, *muslim*, *protestant*. These are the domestic or international determinants to which countries's democracy scores independently respond and have nothing to do with the diffusion mechanism. The second and the third curly brackets contain the possible diffusion processes. The second part is for diffusion through exogenously-shaped connectivities, *border* and *trade volume*. The third part is for the diffusion through endogenously-shaped ties, and this term generates the co-evolutionary dynamic between the outcome variable and the connectivity $\mathbf{L}$, which consists of the pairwise difference of the regie score ($\mathbf{y}$) from the previous time period.

Model (1) is a non-spatial specification and Model (2) is a traditional multiple-spatial-lag model without the co-evolution term. The likelihood-ration (LR) tests find that the difference between Model (1) and (2), and Model (2) and (3) are statistically significant (with the LRs being 21.64 and 8.9 respectively, and the chi-square critical values being 13.82 and 6.63 respectively at the significance level of 1%). It is likely that the non-spatial model (1) suffers from omitted variable bias and it shows up in the overestimated degree of temporal persistency.

My discussion will, therefore, focus on the spatial models (2), (3) and (4). Both Model (3)

Table 4.3: Estimation Results: Regressions of the Unified Democracy Scores on the Country-Attribute Variables and Other States' Democracy Scores

		Beck et al.	OLS	M-STAR	**M-STAR + Co-Evolution**	
		(Cross-sec.)	(1)	(2)	**(3)**	(4)
Common	**Constant**	-13.24***	0.150**	0.099	**0.099**	0.078
exposure		-3.11	(0.063)	(0.067)	(0.069)	(0.080)
($\boldsymbol{\beta}, \phi$)	**Temporal lag**		0.862***	0.826***	**0.826***	0.820***
			(0.015)	(0.016)	(0.016)	(0.016)
	Real GDP/cap		0.016***	0.016***	**0.016***	0.012**
			(0.005)	(0.005)	(0.005)	(0.006)
	(Real GDP/cap)^2		-0.0005***	-0.0005***	**-0.0005***	-0.0004**
			(0.0002)	(0.0002)	(0.0002)	(0.0002)
	Growth rate					-0.004
						(0.003)
	Commonwealth					0.033
						(0.025)
	Urban pop rate					0.001
						(0.0006)
	Fuel export rate		-0.0008***	-0.001***	**-0.0009***	-0.001***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
	Catholic					-0.0001
						(0.0005)
	Muslim					-0.001*
						(0.0004)
	Protestant					-0.0001
						(0.001)
	log(GDP/cap)	1.53***				
		(0.37)				
Exogenous	Geog Distance	0.89***				
connectivity		(0.19)				
($\boldsymbol{\rho}$)	**Borders**			0.053***	**0.052***	0.054***
				(0.019)	(0.019)	(0.019)
	Trade	0.59		0.105***	**0.105***	0.1090***
		(0.43)		(0.027)	(0.027)	(0.028)
Endogenous	**Regime distance**				**-0.104***	-0.098*
connectivity					(0.055)	(0.055)
(γ)						
Other	$\boldsymbol{\sigma}$		0.307***	0.303***	**0.303***	0.302***
parameters			(0.006)	(0.006)	(0.006)	(0.006)
	Time dummies?	NA	Yes	Yes	**Yes**	Yes
	Region dummies?	No	Yes	Yes	**Yes**	Yes
	Log-likelihood		-279.039	-268.219	**-263.769**	-259.601

Standard errors are in parentheses. Significance levels : $*$: 10% $**$: 5% $***$: 1%. Note that Beck, Gleditsch and Beardsley (2006)'s results are based on a cross-sectional specification with the 1998 data. Their dependent variable is from the Polity IV data. The dependent variable of Model (1)-(4) is the Unified Democracy Scores derived in Pemstein, Meserve and Melton (2008). Model (1) is an assumed-independence model, where spatial interdependence is assumed to be zero. Model (2) is a simple M-STAR model with no co-evolutionary dynamic; i.e., all the spatial lags are pre-determined. Model (3) and (4) are M-STAR models with the co-evolutionary dynamic; i.e., these models also include the regime-distance "spatial" lag that is endogenous over time. All the models are estimated with eight temporal dummies and seven regional dummies. The seven regions are Africa, North America, Central and South America, Asia, Middle East, Western Europe and Eastern Europe. The eighth regional dummy for Oceania is dropped from the regression equation.

and (4) utilize the new spatial econometric technique to evaluate the diffusion of democracy allowing for the co-evolutionary dynamic. The only difference between (3) and (4) is that Model (4) contains all the country-specific explanatory variables that I considered for this study, based on substantive theories and following the conventions in the traditional democratization literature. However, the LR test suggests that Model (3) is the preferred specification. Even after adding six more explanatory variables, the LR test cannot reject the null hypothesis that the added parameters in Model (4) are jointly zero. (LR= 8.34, the chi-square critical value with the degrees of freedom 6= 10.65 at the significance level of 10%.) I left Model (4) in the results table to demonstrate the entire list of variables that I consider, but I will use Model (3) to conduct additional analyses in the later sections.

4.10.1 The Effects of Country-Specific Attributes

In interpreting the magnitude of the coefficient estimates, it is important to note that, for most time-periods, the democracy scores of the 75 to 175 countries lie in the range approximately between -2 and 2. For example, in the 1991-1995 time-block ("year 8"), the data include 154 countries. This implies that only about a 0.026-unit increase in the democracy score is necessary, on average, for a country to catch up with the next higher-ranked democracy.

The *real GDP/capita* variables are both highly statistically significant and the signs indicate that there is a curve-linear relationship between the real GDP per capita level and the democracy score where the level of democracy goes up as the income level increases with a diminishing curvature. The result suggests that, for a middle income country, the democracy score could increase by as much as 0.13, as real GDP/capita increases by one unit, or $1000. This is equivalent to say, on average, a middle income country could improve its democracy ranking by 5, all else equal. However, the *growth rate*, as the log-difference of the real GDP/capita levels, turns out not statistically significant. This result confirms the basic findings both in Przeworski et al. (2000) and Boix (2003). Przeworski et al. (2000) points to the asymmetry of the effect of growth in democracies and autocracies. They find that democracies are more sensitive to economic crises (or sudden negative growths) than autocracies. Another possibility, as I mentioned in the previous section, is the existence of multiple competing forces that growth can generate. In countries where the poor can make a credible commitment to maintain moderate tax rates in the potential future democracy, growth could increase the the propensity of democratic transitions, because the poor take the growing economy as an opportunity to receive more redistribution even at an unfair redistributive rate (the rate is low but the pie itself is growing fast enough to make the benefit sufficient) and the elite/rich are convinced that, even after the democratization, the tax they will incur is sufficiently small. At the same time, in autocracies where there is a modest redistributive system, the level of economic grievances could be too low for the repressed to stand against the authoritarian incumbent when the economy grows fast. Without controlling for such institutional differences across countries, it is impossible to entangle the complex effects of growth.

Another country-specific variable that consistently turns out significant is *fuel export rate*. Even though the magnitude is very low, a percent increase in fuel export (out of all the merchandise exports) seems to decrease the democracy score by 0.0009 all else equal. It should be noted that Model (3) also includes a dummy for the Middle East region, which is a typical oil-rich region, and both the region dummy and the fuel export rate are statistically significant, at the 5% and the 1% significance levels respectively. This is strong evidence of the fuel effect or the country-specific-asset effect on democracy, confirming one of Boix (2003)'s main claims.

All the other country-specific variables turn out to be statistically insignificant. The results are very robust in that regardless of the combination of these common-exposure variables, only the coefficients of *temporal lag*, *real GDP/capita*, *(real GDP/capita)*^{2} and *fuel export rate* are statistically significant most of the time and not others. These results are not necessarily surprising. Przeworski et al. (2000) also find weak or no results of the religion dummies. In my Model (4), the *muslim* variable has a modestly significant and negative effect on democracy, but the Middle East region dummy in this specification becomes not significant while it is almost always significant in other specifications. From this, there is no way to tell empirically whether it is something about the Middle East region in general that prevents democratization, or whether there is something particular about their culture related to the religion. Also the high correlation between urbanization and economic development might be the reason why the coefficient for *urban population rate* is not statistically different from zero.

Finally, all the specifications find strong significant and positive effects of the temporal lag, confirming that history of each country's regime matters and the magnitude is large.

4.10.2 Diffusion: the Effects of Other Countries' Democracy Levels

Now let us look at the estimated coefficients related to the diffusion and self-selection. The estimates for the exogenous-diffusion terms (ρ's), in all three spatial models, uncover positive interdependence of political regimes through the geographical and economic ties. Note the the strength of the effects of other countries' democracy scores through a particular connectivity can be computed as $\hat{\rho}_r \mathbf{W}_r$; i.e., the information we can obtain from the coefficient ρ's themselves is merely to what extent the connectivity of units matters overall, and it is not the magnitude of "influence" that other countries' regimes have. I will later compute the overall strength of other's influence in the next section. What we can learn from the significant estimates of *border* and *trade* is that to whom others and to what extent a country is connected matter in predicting a country's democracy level.

Finally the main contribution of this paper, the influence of other regimes through regime distance, is uncovered by the estimate of γ. The coefficients are negative and statistically significant. Note that the connectivity matrix $\mathbf{L}$ carries the information about the "magnitude" of the pairwise regime dissimilarity; i.e., all the entries are absolute values of the distance between political regimes of any given two countries. Since all the entries of $\mathbf{L}$ are greater than or equal to 0, the negative estimate of γ combined with the positive estimate for ρ_r's implies that a country's political regime influences those of countries with more similar regimes than dissimilar regimes: homophily in other words. This could imply that countries with similar democracy levels become more similar over time as if they were reinforcing each other's regime type, possibly generating several regime "blocs" in the long run. It should be noted that these regime blocs/clubs are not necessarily clustered geographically, if the effect of regime-similarity-based contagion is substantial compared to the effect of geography-based contagion, for example. I will discuss the long-term distribution of political regimes around the world in the later section on counterfactual simulations.

I have developed this somewhat unified model of the diffusion of democracy, building on the accumulated knowledge about the determinants of democratization in the existing studies. From the coefficient estimates of the most preferred specification, (3), we can conclude that countries' regime scores are highly correlated with some country-specific factors, such as their economic status and the asset specificity. After controlling for the effects of such variables, I find empirical evidence that there is true interdependence in countries' democracy levels and that geographical contiguity and trade volumes are at least partially defining the strength of this interdependence. Moreover, as I suspected, the democracy scores achieved by these countries in a certain time period partially determine the degrees of regime influence among them in the next time period. Even though it is difficult to see the magnitude of the selection bias in the coefficient estimates, it is suggestive that what we thought before was the diffusion of regimes occurring through exogenous ties is in fact partly due to the fact that countries with similar regime scores are more likely to have strong ties precisely because of their similarity in regimes scores. In other words, countries are selecting themselves into regime networks that in turn shape their next-stage democracy levels. Before the M-STAR model with co-evolutionary dynamics, we could not test for the possibility of this selection.

4.10.3 Discussion of Estimation Results

Do countries with similar levels of democracy influence on each other's regimes more strongly in the future? To put it more abstractly, does a country form stronger ties in the future with countries with more similar regime scores than different scores? This was the main empirical question asked in this project. In the previous section, I estimated coefficients based on a model that distinguishes the possibility of countries' endogenous tie-forming behavior (endogenous network formation) from the effects of other's democracy levels through already-defined country networks (traditional diffusion of democracy). The analysis showed that there is indeed a statistically significant dynamic of endogenous tie forming. In other words, countries that manifested similar democracy levels in the past form stronger (invisible) ties, and hence influence each other's democracy level more powerfully in the next time period.

As a results, the following dynamic emerges over time: more democratic countries influence each other such that it reinforces their democratic regimes, and similarly more autocratic countries influence each other's regime such that it reinforce their autocratic regimes over time. The first counterfactual analysis in the following demonstrates the behavior-reinforcing dynamic by an example, based on the model I used for estimation. It shows, by simulation, that th effects of a democratic shock to a country get inflated much more and much faster in a model that incorporates co-evolution of networks and democracy levels. Given that the empirical analysis shows there is indeed a statistically significant co-evolutionary dynamic in the process of democratization, this means that the prediction of democracy levels based on traditional diffusion models can be very misleading.

Furthermore, from the fact that the very outcome of the model—the level of democracy—

contributes to countries' tie-forming mechanism in the future, it is reasonable to suspect that a country's intertemporal regime trajectory depends on "when" the country receives an external democratic or autocratic shock. By "when", I mean (i) the structure of connectivity among the countries (pre-defined networks) and (ii) the level of democracy of the county itself and of all the other countries'. It might be more intuitive to imagine a profile of all the countries' democracy scores around the world. The bottom line is; a country's intertemporal regime trajectory should depend on the political environment in which the country receive a democratic shock, and not only the magnitude of democratic it receives. For this reason, I suspect that the process of democratization is path dependent, or at least history dependent.[23] Since the spatial econometric model used in my empirical analysis is not directly based on a theoretical model that demonstrates path dependency, I cannot prove rigorously that the data generating process in democratization is path dependent in a strict mathematical sense. However, in the second counterfactual analysis, I will show graphically that it is indicative that democratization is history dependent; i.e., the development of regimes is sensitive to initial conditions of the unit itself, and the surrounding environment.

4.11 Long-Run Implications of the Regime-Support (Self-Selection) Network

4.11.1 Simulation 1: Regime Type Can Be Reinforcing Over Time

For each set of observed or counterfactual data, there is a set of steady-state (or long-run) levels of democracy to which countries eventually reach assuming that the observed or counterfactual values for each variable won't change over time. Clearly, time-invariance of the

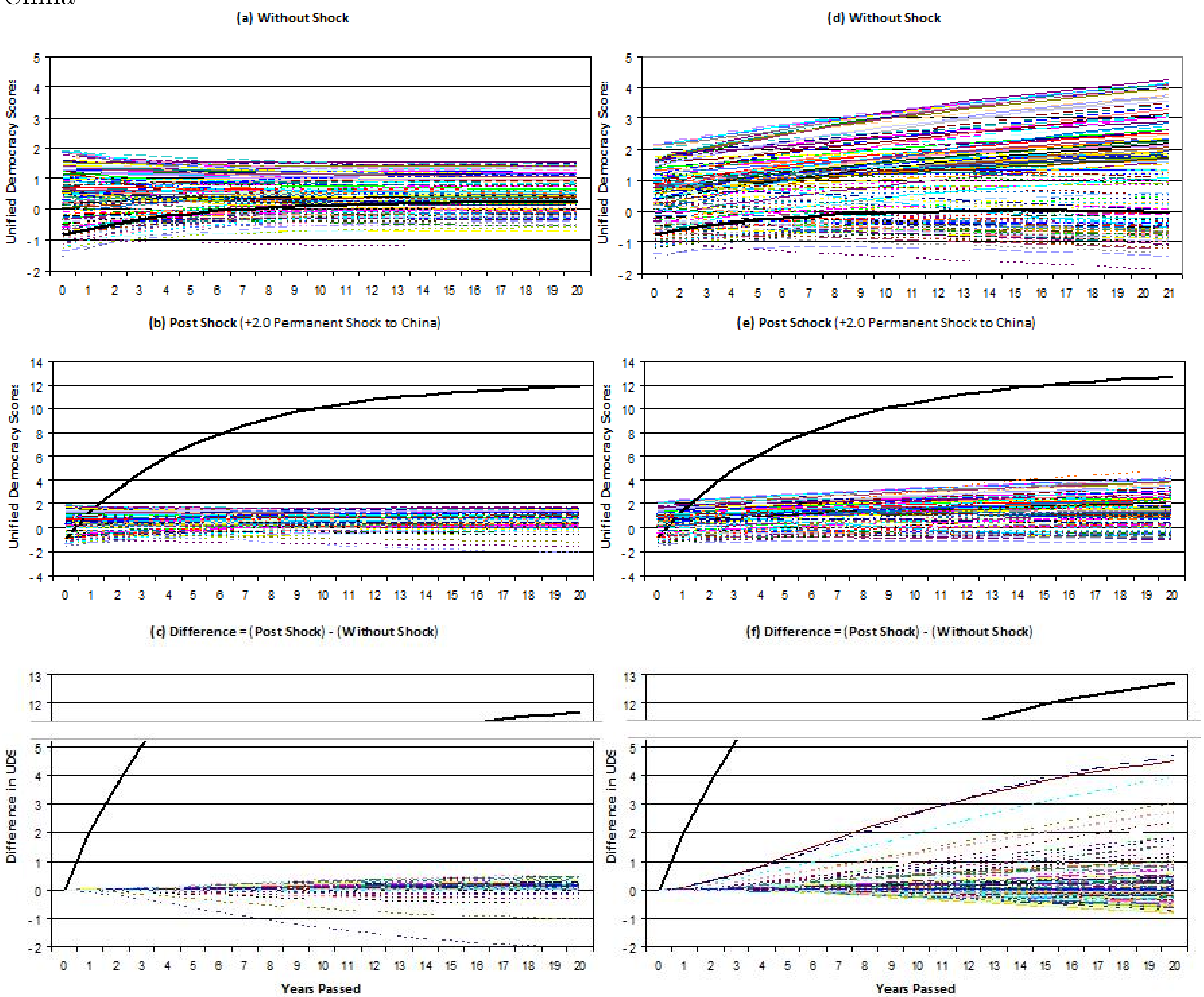
[23] Cite Page 2006 and explain briefly what they mean.

variables is not a realistic assumption, but it is still useful to conduct some counterfactual analyses. For example, it is difficult for us to foresee how a democratic change in one country spreads across the world through the estimated interdependence until we simulate such changes. Or a promoter of democracy might want to know how an improvement in the average income level in one country affects the country's own *and* other countries' democracy levels in the following years. Obviously this is where theories about diffusion matter.

As an example, I gave a substantial democratic shock to China (+2.0 in the UDS unit) and observed how it changed the trajectories of all the countries' democracy levels over the following 20 years. For the spatial weights and the other variables, I used the values from the last time period (1996-2000), except the dependent variable which was computed using all the other variables and the estimated coefficients (i.e., the starting levels of the democracy scores are the fitted values). If the system maintain those variables' values in the following years, the countries' democracy scores follow the trajectories plotted in Figure 4.9-(d) (the top-right panel of Figure 4.9). Eventually all the countries would reach their own steady-state levels after many iterations, but I limited the length of simulation to 20 years (20 iterations), because an extreme extrapolation would not be a good empirical practice and the values for the explanatory variables fixed at the 1996-2000 level would become less and less realistic as time passes.

With a +2.0 shock to China's democracy level, the trajectories look like the ones in panel (e). The curves for China are highlighted in black and bold in all the graphs. The difference between the with- and without-shock curves can be seen more clearly in the last panel, (f). Graph (f) plots the difference of the with- and without-shock regime trajectory: as can be seen in the graph, the difference in the democracy score keeps increasing for some countries, but decreasing or remains the same for others. If the difference increases over time, that means that the positive shock to China gave a positive influence to the country's democracy score and the (positive) gap between the original (no-shock) change in its regime

Figure 4.9: Trajectories of Democracy Scores with and without a Counterfactual Shock in China



score accelerates over time. China, obviously, as well as Maldives and North Korea follow this pattern to a great extent.

The growing gap between the original trajectory and the trajectory after a shock mainly stems from the effect of co-evolution. When the shock occurs, a country's democracy level jumps up or down depending on its country-specific characteristics and through the estimated interdependence paths. Now with the new level of democracy in these countries, the new connectivity is defined both by the similarity in the new regime scores and the exogenous connectivities (i.e., trade and border). The new network now determines the level of these countries' democracy scores in the next iteration together with their country-specific variables. The co-evolutionary dynamic generated by the evolving similarity connectivity reinforces the direction of the regime change for most of the countries, whether upward or downward. This is mainly why we observe amplified effects of a shock over time.

The effect of co-evolutionary dynamic becomes much more obvious when we compare panel (f) with panel (c), which is the same difference but generated from a specification that does not have the similarity (endogenous) connectivity. I used Model (2) to conduct the simulations for panel (a), (b) and (c). Comparing (c) and (f), it is clear that we could underestimate the effect of shocks if the actual process in democracies is closer to the specification with co-evolution but we estimate the effect using the traditional spatial models, which only contain exogenous connectivities.

4.11.2 Simulation 2: Indication of History Dependency in Democratization

In this section, I demonstrate using simulations that the paths of the outcome values toward their steady-states can vary dramatically depending on the initial outcome values of all the units, and this is the case *only when the self-selective country ties are included in* *the model*. More precisely, and in the context of democratization, when only conventional contagion processes are included (i.e., there are no self-selective dependency networks like regime support), (1) democracy levels of *all the countries* countries converge to a single level in the long run regardless of the initial/earlier democracy levels of these countries, and (2) they converge to exactly the same level of democracy regardless of the initial/earlier regime levels of these countries. On the contrary, when a self-selective (such as the regime-support network) dependency network is present, the long-run trajectory of the system looks very different: (1) countries' democracy levels can end up at different values depending on the initial democracy profile of the world, and (2) the set of democracy levels at which countries end up can be different depending on the initial democracy profile of the world. This is an astonishing finding, at least from the substantive perspective. In other words, these simulations suggest that spatial models used for the conventional contagion mechanisms assume away a system dynamic in which outcome levels carry the influence from their values earlier in the history.

To demonstrate clearly this difference in two system dynamics, I simulate the evolution of the outcome values using the following very simple system with only 4 units. This is equivalent to model a world of only four countries. Equation 4.12 represents the model at a single time period t.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \end{bmatrix} = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \\ x_{4,t} \end{bmatrix} + \phi \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \end{bmatrix} + \rho \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \end{bmatrix}$$

$$+ \gamma \begin{bmatrix} 0 & |y_{1,t-1} - y_{2,t-1}| & |y_{1,t-1} - y_{3,t-1}| & |y_{1,t-1} - y_{4,t-1}| \\ |y_{2,t-1} - y_{1,t-1}| & 0 & |y_{2,t-1} - y_{3,t-1}| & |y_{2,t-1} - y_{4,t-1}| \\ |y_{3,t-1} - y_{1,t-1}| & |y_{3,t-1} - y_{2,t-1}| & 0 & |y_{3,t-1} - y_{4,t-1}| \\ |y_{4,t-1} - y_{1,t-1}| & |y_{4,t-1} - y_{2,t-1}| & |y_{4,t-1} - y_{3,t-1}| & 0 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \end{bmatrix} \tag{4.12}$$

The term on the left-hand side (LHS) is a vector of outcome levels for the four observational units $i = 1, \cdots 4$. In my study, this is a vector of democracy levels. n the right-had side (RHS), the first term is a vector that represents typical covariates included in models for democracy levels. In the real world, these include country-specific characteristics such as the economic development level, resource-richness, religious profile and so forth. In this simple simulation model, all these unit-specific attributes are summarized in a single $x_{i,t}$ without loss of generality (WLOG). The second term is a unit's own outcome level from the precious time period, representing a typical "time-lag" variable. For example, countries that were highly democratic in the previous time period are relatively more likely to maintain higher democracy scores in the next period. The third term captures conventional contagion through non-selective dependency networks such as geographical proximity and trade flows. WLOG, I include a single dependency network, and assume the connectivity across the four units is time-invariant. I also assume the connectivity is dichotomous—either connected or not. Countries are connected expecting some regime influence between them, if the cell value is 1. If the value is 0, there is no connection, and hence, no directly contagion between the two.[24] Finally the last term represents contagion through a self-selective network, such as

[24]But remember, indirect contagion is still possible between countries that does not directly share a

regime support. See how each cell of the contiguity matrix is defined by the absolute difference between two units' past outcome levels. This contiguity matrix, therefore, measures how different each pair of units were in the last time period, in terms of the outcome values.[25] Contagion of regimes can occur both through the non-selective and selective ties in the third and fourth terms of the equation.

I give a fixed set of parameter values as follows;

- $\phi = 0.5$
- $\rho = 0.4$
- $\gamma = -0.5$

And the covariates as follows;

- $\begin{pmatrix} x_{1,1} \\ x_{2,1} \\ x_{3,1} \\ x_{4,1} \end{pmatrix} = \begin{pmatrix} -5 \\ 0 \\ 10 \\ 5 \end{pmatrix}, \begin{pmatrix} x_{1,2} \\ x_{2,2} \\ x_{3,2} \\ x_{4,2} \end{pmatrix} = \begin{pmatrix} 0 \\ -5 \\ 10 \\ 0 \end{pmatrix}, \begin{pmatrix} x_{1,3...} \\ x_{2,3...} \\ x_{3,3...} \\ x_{4,3...} \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \\ 10 \\ 10 \end{pmatrix}$

It was not necessary to change the covariate values for the earlier time periods as I did here (the x values for time period 1 an 2), but I decided to set up the simulation this way, so that it will be useful in the future, when I study the system characteristics in terms of *path dependency* in a mathematically rigorous framework.[26]

The key instruments for this simulation analysis is the initial outcome values fed into the system. Sets of numbers associated with each graph in Figure 4.10 and Figure 4.11 are these initial conditions. I consider two systems (models) in these simulations. First, a model without the homophilic network, or the fourth term of Equation 4.12. The second model is with the homophilic contagion path, exactly the same form as in Equation 4.12. I let both

network tie, due to feedback.

[25] Strictly speaking, this is heterophily instead of homophily, but we can switch our language in interpreting the results by flipping the sign.

[26] I discuss part of this issue in Appendix A.3.2, a research note for a future study.

models run with the aforementioned parameter and covariate values and each of the three different set of initial outcome values. The three distinct sets of initial outcome values are (1) $\{20, 0, 0, -20\}$, (2) $\{10, 0, 0, -10\}$, and (3) $\{0, 10, 10, 0\}$.

Figure 4.10: Long-Run Evolution of the Dependent-Variable Values with Traditional Spatial Models

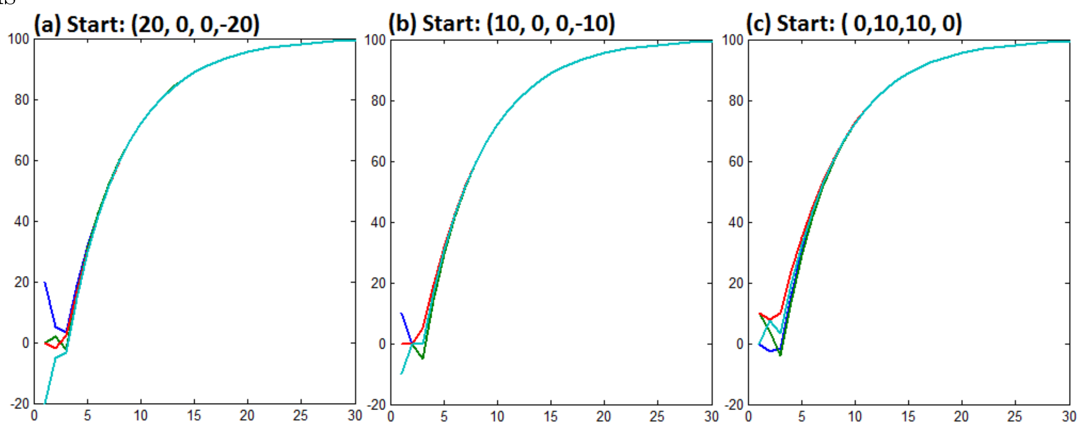


Figure 4.11: Long-Run Evolution of the Dependent-Variable Values with an M-STAR+Co-Evolution Model

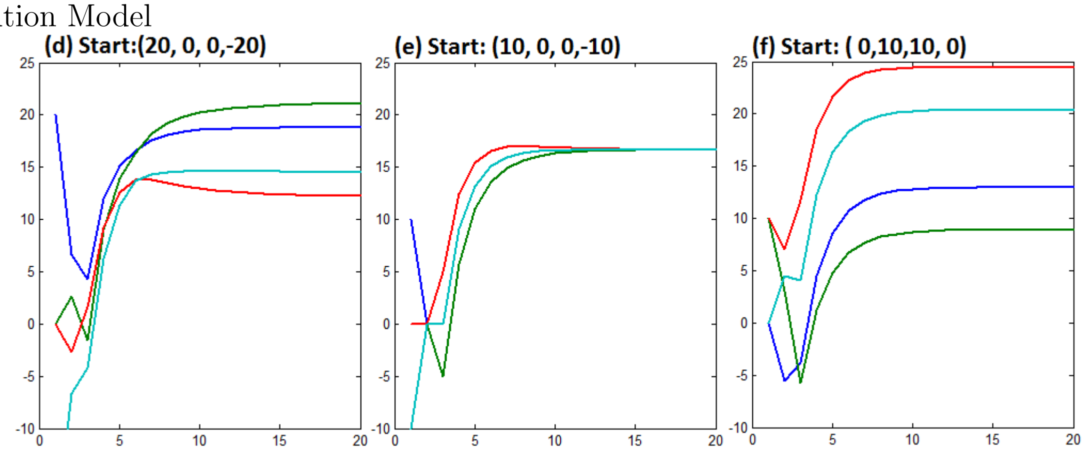


Figure 4.10 presents the results of simulations *without* contagion by homopilic networks. Figure 4.11 presents the results of simulations *with* contagion by homopilic networks. The difference is stark. In the model without reinforcement, outcomes of all the units converge to a single level, and they converge to exactly the same level regardless of the initial conditions, while in the reinforcement model, one demonstrates a global convergence and the other two demonstrate global divergences, each to different levels. Even more interestingly, these two

diverging cases also demonstrate two clusterings. The implication of this result is particularly noteworthy, because if I apply this finding to the context of democratization, the conventional spatial approach (i.e., without selection) would be assuming, just by the choice of the model, that all the states' democracy levels should converge, and would converge to exactly the same level of democracy at some point in the future, *regardless of* the distribution and topology of surrounding political regimes; i.e., the initial democracy profile of the world is irrelevant to a country's regime trajectory in the long run. This is a very unlikely scenario in the real world.

4.12 Conclusion

This study revisited the meaning of diffusion in regime transitions and proposed a type of diffusion path that has been overlooked in existing studies. I posit a theory that a country form stronger dependency ties with countries that demonstrated similar democracy levels in the past (homophily). Regime support or approval was the rationale for such a self-selective network formation among countries. Incorporating countries' self-selection generates a co-evolutionary dynamic between the dependent variable (behavioral type) and a dependence network among countries. I introduced a new spatial estimator, "MSTAR + Co-evolution" (MSTARC) model, in order to distinguish three processes that can determine a country's democracy level. The first process is where country-specific attributes determine the level of democracy as typically suggested in the earliest democratization studies. The second process is the traditional diffusion mechanism in which other states' democracy levels predict a country's regime. Finally the third type of process is my theory of regime reinforcement. This third process is important to consider because the seeming diffusion effect (the second mechanism) can be partially a mere consequence of countries' self-selection into peer regime-support networks, potentially inflating the estimated effect of diffusion in the existing empirical studies. The empirical analyses find that other states' democracy levels indeed affect a country "positively" through the self-selective (or the regime-similarity) connectivity.

The first implication of this finding is that a selection mechanism (homophily) exists in political regimes transitions. Part of the estimated influence of others' regimes that travels through the exogenous paths (trade and border) in existing studies could be overestimated due to the fact that countries self-select themselves into regime networks.

Another important implication of the co-evolutionary dynamic becomes evident in the simulation analyses. Since selection in democracy levels reinforces the current direction of change (i.e., democratizing or autocratizing) in each country, this non-linearity in the system becomes sensitive to initial conditions. The initial conditions influence the path that a country's democracy score follows later. Since this dynamic was not a part of traditional models, we were not able to test for path dependency before. In fact, by using the tradition models for diffusion, one would be implicitly assuming that regardless of the initial conditions (or a set of conditions surrounding country A at one time period), country A would always reach its own steady-state level, x, if the system runs for a long time. For example, which countries in the world have more democratic regimes in a certain year does not matter to one's long-run democracy level in this framework. However, it should be intuitive for any social scientist that two worlds with different distributions of regimes have very different implications for the political development of a country. Introducing the co-evolutionary dynamic enables us to address this issue.

Lastly another implication that has not been discussed in this paper throughly (yet) is the exact working of path-dependence. As I mentioned earlier, the path-dependence generated by the homophilic network-formation over time reinforces the current direction of changes in the democracy scores. If this is the case, then there should be a border line/lines above which countries head toward more democratic regimes and below which countries head to more autocratic regimes. Eventually there should be several "convergence clubs". Under what conditions should we expect to observe a world with a complete convergence, dichotomous regime clubs, trichotomous clubs, and so on? And why do we see a two-regime world currently? To answer these questions, I will need to further explore the workings of the co-evolutionary dynamic both methodologically and substantively. I discuss a possible research design in Appendix A.3.2

Appendix A

Appendix

A.1 Appendices for Essay 1

A.1.1 Mathematical Properties of the Bivariate Weibull Distribution

In the following, we briefly describe some mathematical properties of the bivariate Weibull distribution presented above. Although it was already proven that any bivariate distribution that belongs to the Farlie-Gumbel-Morgenstern family satisfies the axioms of probability (Gumbel 1959, 1960), the following mathematical properties that lead to the derivation of ρ are useful.

As shown in Gumbel (1959) and Gumbel (1960), a bivariate distribution function can be constructed from two marginal probability functions, $F(y_1)$ and $F(y_2)$, by the following copula with a constraint on parameter α, $-1 \leq \alpha \leq 1$;

$$F(y_1, y_2) = F(y_1)F(y_2)[1 + \alpha\{1 - F(y_1)\}\{1 - F(y_2)\}], \tag{A.1}$$

where $-1 \leq \alpha \leq 1$.

The associated joint density function is given as follows;

$$f(y_1, y_2) = f(y_1)f(y_2)[1 + \alpha\{2F(y_1) - 1\}\{2F(y_2) - 1\}]. \tag{A.2}$$

Accordingly, a bivariate Weibull distribution can be constructed from the following marginal distributions and densities;

$$F(y_i) = 1 - e^{-(\frac{y_i}{\theta_i})^{\lambda_i}},$$

$$f(y_i) = \frac{\lambda_i}{\theta_i}(\frac{y_i}{\theta_i})^{\lambda_i - 1} e^{-(\frac{y_i}{\theta_i})^{\lambda_i}}; i = 1, 2,$$

where $\lambda_i > 0$ and $\theta_i > 0$.

The joint probability and density are given by

$$F(y_1, y_2) = (1 - e^{-(\frac{y_1}{\theta_1})^{\lambda_1}})(1 - e^{-(\frac{y_2}{\theta_2})^{\lambda_2}})(1 + \alpha e^{-(\frac{y_1}{\theta_1})^{\lambda_1} - (\frac{y_2}{\theta_2})^{\lambda_2}}) \tag{A.3}$$

$$f(y_1, y_2) = \frac{\lambda_1}{\theta_1}\frac{\lambda_2}{\theta_2}(\frac{y_1}{\theta_1})^{\lambda_1 - 1}(\frac{y_2}{\theta_2})^{\lambda_2 - 1} e^{-2[(\frac{y_1}{\theta_1})^{\lambda_1} + (\frac{y_2}{\theta_2})^{\lambda_2}]}[4\alpha - 2\alpha e^{(\frac{y_1}{\theta_1})^{\lambda_1}} - 2\alpha e^{(\frac{y_2}{\theta_2})^{\lambda_2}} + (1+\alpha) e^{(\frac{y_1}{\theta_1})^{\lambda_1} + (\frac{y_2}{\theta_2})^{\lambda_2}}], \tag{A.4}$$

where $y_i \geq 0$, $-1 \leq \alpha \leq 1$, $\theta_i > 0$ and $\lambda_i \geq 0$.

The joint cdf must satisfy the following boundary conditions

$$\begin{cases} F(0, y_2) = F(y_1, 0) = 0 \\ F(\infty, \infty) = 1, \end{cases}$$

and the joint density has to be nonnegative, $f(y_1, y_2) \geq 0$. Any bivariate distribution that belongs to the Farlie-Gumbel-Morgenstern family satisfies these conditions.

Another condition a bivariate distribution always has to satisfy is Fréchet's inequality,

$$F(y_1, y_2) \leq F_1(y_i); i = 1, 2 \tag{A.5}$$

for all y_1 and y_2. Since the dependence of the joint distribution and density on y_1 and y_2 is symmetric, it is sufficient to explore the performance of one variable y_1. This applies to all the calculations in the rest of this appendix. From (A.3) it follows after a simplification that

$$\alpha e^{-(\frac{y_1}{\theta_1})^{\lambda_1}}(1 - e^{-(\frac{y_2}{\theta_2})^{\lambda_2}}) \leq 1. \tag{A.6}$$

In sum, the function (A.3) satisfies all the required axioms of probability function, under the conditions of $-1 \leq \alpha \leq 1$, $\theta_i > 0$ and $\lambda_i \geq 0$.

The followings are the relevant computations to derive the correlation coefficient ρ. By definition, the correlation coefficient of two random variables, Y_1 and Y_2 can be obtained as

$$\rho = \frac{E(y_1 y_2) - E(y_1)E(y_2)}{\sigma_{y_1}\sigma_{y_2}}. \tag{A.7}$$

For our marginal probabilities, $F(y_1)$ and $F(y_2)$, the means and variances are

$$\begin{aligned} E(y_i) &= \theta_i \Gamma(1 + \frac{1}{\lambda_i}) = \theta_i \frac{1}{\lambda_i}\Gamma(\frac{1}{\lambda_i}) \\ Var(y_i) &= \theta_i^{\frac{2}{\lambda_i}}[\Gamma(1 + \frac{2}{\lambda_i}) - \Gamma^2(1 + \frac{1}{\lambda_i})]; \quad i = 1, 2. \end{aligned} \tag{A.8}$$

Now the only term we need to compute to obtain ρ is $E(y_1 y_2)$. From (15), the marginal densities are

$$f(y_i) = \int_0^\infty f(y_1, y_2) dy_i = \frac{\lambda_i}{\theta_i}(\frac{y_i}{\theta_i})^{\lambda_i - 1} e^{-(\frac{y_i}{\theta_i})^{\lambda_i}}; i = 1, 2, \tag{A.9}$$

which is, of course, the Weibull densities.

The conditional expectation of y_1 can be obtained as

$$\begin{aligned} E(y_1|y_2) &= \int_0^\infty y_1 f(y_1|y_2) dy_1 \\ &= \frac{1}{\lambda_1}\theta_1 \Gamma(\frac{1}{\lambda_1}) 2^{-\frac{1}{\lambda_1}} e^{-(\frac{y_2}{\theta_2})^{\lambda_2}}[-2\alpha(2^{\frac{1}{\lambda_1}} - 1) + e^{(\frac{y_2}{\theta_2})^{\lambda_2}}(2^{\frac{1}{\lambda_1}}(1 + \alpha) - \alpha)], \end{aligned} \tag{A.10}$$

where

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f(y_2)}. \tag{A.11}$$

The expectation of the cross-product can be computed as

$$\begin{aligned} E(y_1 y_2) &= \int_0^\infty y_2 E(y_1|y_2) f(y_2) dy_2 \\ &= \frac{\theta_1}{\lambda_1}\frac{\theta_2}{\lambda_2}\Gamma(\frac{1}{\lambda_1})\Gamma(\frac{2}{\lambda_2})\left[1 + \alpha(1 - 2^{-\frac{1}{\lambda_1}} - 2^{-\frac{1}{\lambda_2}} + 2^{-\frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}})\right]. \end{aligned} \tag{A.12}$$

Substituting (A.12) into (A.7), we get

$$\rho = \frac{2^{-\frac{\lambda 1+\lambda 2}{\lambda 1 \lambda 2}}\left(2^{\frac{1}{\lambda 1}} - 1\right)\left(2^{\frac{1}{\lambda 2}} - 1\right)\alpha\Gamma\left[\frac{1}{\lambda 1}\right]\Gamma\left[\frac{1}{\lambda 2}\right]}{\lambda 1 \lambda 2\sqrt{-\Gamma^2\left[1 + \frac{1}{\lambda 1}\right] + \Gamma\left[\frac{2+\lambda 1}{\lambda 1}\right]}\sqrt{-\Gamma^2\left[1 + \frac{1}{\lambda 2}\right] + \Gamma\left[\frac{2+\lambda 2}{\lambda 2}\right]}}. \tag{A.13}$$

Note that the scale parameter θ does not affect the dependence of y_1 and y_2, ρ. As mentioned in the previous sections, ρ is increasing in α and the maximum range is $-.322409 \leq \rho \leq .322409$ when $\lambda_1 = \lambda_2 = 3.29035$.

A.1.2 The Logged Weibull and the Standard Gumbel Variables

In the derivation of the Weibull FIML estimator, I claimed that a logged Weibull random variable is a Standard-Gumbel –a special case of the type-I extreme value (minimum)– variable that is scaled by the inverse of the original Weibull shape parameter. This section demonstrates the log transformation of a Weibull variable. Recall the density and distribution functions of the Standard Gumbel distribution and the Weibull distribution;

$$\text{Standard Gumbel distribution} \begin{cases} f(u) = e^{u} e^{-e^{u}} \\ F(u) = 1 - e^{-e^{u}}, \end{cases}$$

$$\text{Weibull} \begin{cases} f(y) = \frac{\lambda}{\theta}(\frac{y}{\theta})^{\lambda-1} e^{-(\frac{y}{\theta})^{\lambda}} \\ F(y) = 1 - e^{-(\frac{y}{\theta})^{\lambda}}, \end{cases}$$

where θ is a scale parameter and λ is a shape parameter.

Consider a Weibull random variable Y that is scaled by θ. The log of the Weibull variable is a Standard Gumbel variable, U, scaled by the inverse of the Weibull shape parameter, $\frac{1}{\lambda}$. If this statement is true, then the following holds;

$$\begin{aligned} \frac{1}{\lambda} u &= \ln\left(\frac{y}{\theta}\right) \\ &\Leftrightarrow y = \theta e^{\frac{u}{\lambda}}. \end{aligned}$$

Since $Y \sim Weibull(\lambda, \theta)$,

$$\begin{aligned} F(y) &= 1 - e^{-(\frac{y}{\theta})^{\lambda}} \\ &= 1 - e^{-(\frac{\theta e^{\frac{u}{\lambda}}}{\theta})^{\lambda}} \\ &= 1 - e^{-e^{u}} = G(u). \end{aligned} \tag{A.14}$$

$G(u)$ is the cdf of the Standard Gumbel distribution.

The moments for this extreme value distribution are given as follows;

$$E[u] = \gamma,$$

where γ is the Euler-Mascheroni constant, and

$$Var(u) = \frac{\pi^2}{6}.$$

A.1.3 Stata Code Used in the Application

The following is the Stata code used to produce the main results with the interdependent duration model in Table 2.3.

```
***********************************************************
* Program to estimate SDEQ (Weibull) Model
***********************************************************
* Last updated July 14, 2012, by Aya Kachi (Stata Ver.9)
* Original program written by Jude Hays (June 2008)
***********************************************************
clear
pr drop _all
set more off

***********************************************************
* Likelihood evaluator
*----------------------------------------------------------
* Model with two dependent durations
***********************************************************
program define seq_dur_ll
args lnf mu1 mu2 alpha1 alpha2 lambda1 lambda2
tempvar J ay1 ay2
scalar a1 = `alpha1'
scalar a2 = `alpha2'
gen `ay2' = `alpha1'*$ML_y2
gen `ay1' = `alpha2'*$ML_y1
matrix IA = [1, -(a1) \ -(a2), 1]
scalar l1 = `lambda1'
scalar l2 = `lambda2'
matrix L = [l1, 0 \ 0, l2]
matrix IAL = IA*L
qui gen double `J' = ln(det(IAL))
scalar J = `J'

qui replace `lnf'= J + `lambda1'*($ML_y1-`ay2'-`mu1') - exp(`lambda1'*($ML_y1-`ay2'-`mu1')) + /*
*/ `lambda2'*($ML_y2-`ay1'-`mu2') - exp(`lambda2'*($ML_y2-`ay1'-`mu2'))
end

***********************************************************
* Load data
***********************************************************
drop _all
use
"PATH TO THE DATA", clear

* for analysis without mauritius
*drop if survival == 34

gen gdppc_sq_lib = rgdp96pc_gled_100000_lib^2
gen gdppc_sq_rev = rgdp96pc_gled_100000_rev^2

gen lnform = ln(transition)
gen lndur = ln(survival)

global Y1 lnform
global Y2 lndur
global X1 rgdp96pc_gled_100000_lib gdppc_sq_lib commonwealth_2007_lib urbanpercent_10percent_lib dic_military /*
*/ fuelexp_10percent_lib moslem_pacl_10percent_lib
global X2 rgdp96pc_gled_100000_rev gdppc_sq_rev commonwealth_2007_rev urbanpercent_10percent_rev dic_military /*
*/ dic_noindep president moslem_pacl_10percent_rev
*gdppc_sq_rev

***********************************************************
```

```
* Produce starting values by a univariate Weibull duration
* model for each duration equation
**********************************************************
stset transition
streg $X1, dist(weibull) time
matrix stregbp1=e(b)
local col1 = colsof(stregbp1)
matrix stregb1=stregbp1[1,1..‘col1’-1]
matrix coleq stregb1 = mu1
local stregp1=exp(stregbp1[1,‘col1’])
stset survival
streg $X2, dist(weibull) time
matrix stregbp2=e(b)
local col2 = colsof(stregbp2)
matrix stregb2=stregbp2[1,1..‘col2’-1]
matrix coleq stregb2 = mu2
local stregp2=exp(stregbp2[1,‘col2’])

**********************************************************
* Estimate SDEQ (Weibull) model
**********************************************************
ml model lf seq_dur_ll (mu1: $Y1=$X1) (mu2: $Y2=$X2) (alpha1:) (alpha2:) (lambda1:) (lambda2:)
ml init stregb1
ml init stregb2
ml init alpha1:_cons=0
ml init alpha2:_cons=0
ml init lambda1:_cons=‘stregp1’
ml init lambda2:_cons=‘stregp2’
ml max

* to compute aic and bic
gen double aic = -2*e(ll)+2*e(rank)
gen double bic = -2*e(ll) + log(32) * e(rank)
```

A.1.4 Other Applications of Interdependent Duration Models

I present two other applications using interdependent duration models developed in this essay. These application results are also included in a working paper elsewhere, (Hays and Kachi 2011). In the main part of the essay, I focused on the democracy study in Africa because I am interested in studying interdependence in the studies of democratization throughout this thesis, but the two more examples I list here might be helpful for readers to get better ideas of where the statistical model can be applied.

The first example examines the dependence between the formation and dissolution of government in sixteen Western European countries. This is a case where durations of two political "events" (occurring to various actors) are dependent on each other, exemplifying the same dependency structure as the one I studied in this essay using the example of democratization in Africa. The second example here examines the timing of issue position taking in Congress, extending an important work conducted by Box-Steffensmeier, Arnold and Zorn (1997), Boehmke (2006) and Darmofal (2009). The duration of interest is a spell of time until members of Congress announce their issue position on NAFTA. This is a case where a single type of event (time till issue position taking) across different actors are interdependent. Boehmke (2006) first extends the study in Box-Steffensmeier, Arnold and Zorn (1997) by incorporating duration the dependence (but with a "nuisance" approach), and Darmofal (2009) extends the original work by incorporating the spatial dependency across decisions by the members of Congress. We introduce both sources of interdependence.

Interdependence Across Durations: Government Formation and Tenure Durations

Scholarly interest in the empirical determinants of government formation and dissolution in parliamentary democracies is longstanding, and these topics remain among the most central in the comparative study of developed democracies. Two of the most popular topics in this literature are explaining the durations of both coalition bargaining over ministerial portfolios and government survival. There is good reason for this focus. The failure of parliamentary parties to form governments quickly (e.g., the recent crisis in Belgium) and chronic government instability (e.g., Italy for much of the postwar period) have significant social costs and are viewed as symptoms of dysfunctional democracy.

The quantitative empirical literature in this area is large.[1] Typically, the empirical studies explore a set of contextual and cabinet specific factors that determine both kinds of durations. The effects are estimated separately (e.g., for the coalition bargaining duration, Diermeier and van Roozendaal 1998; Martin and Stevenson 2003; and for the cabinet survival, Warwick and Easton 1992; Alt and King 1994, and Diermeier and Stevenson 1999). This is not to

[1]See reviews in Laver (1998, 2003) for more extensive treatments of both literatures.

say that the interdependence has been completely ignored.King et al. (1990) and Warwick (1992), among others, put government formation duration, what they call crisis duration, and the number of formation attempts on the right-hand side of their government survival models.

In the more theoretically oriented literature, Strøm, Budge and Laver (1994) highlight the importance of cabinet termination and dissolution rules for government formation. Fearon (1998) also formalizes the effects of expected enforcement levels of bargained outcomes on the bargaining stage itself, in the context of international agreements. His formulation suggests that a longer shadow of the future can give states an incentive to bargain harder, delaying agreement in hope of getting a better deal. Diermeier, Eraslan and Merlo (2003) also formalize explicitly the interdependence of government formation bargaining and the bargained outcome -cabinet survival. The main purpose of Diermeier, Eraslan and Merlo (2003) is to analyze the conditions under which certain types of coalitions are formed. As an empirical matter, their interest lies in estimating the probability that a particular type of coalition is chosen. Durations of bargaining and government survival still play important roles in their model, but those durations are not the primary focus of their analysis. In their model, the inefficient delay of bargaining is generated mainly by a stochastic factor, the state of the world that is either favorable or unfavorable for a cabinet's survival, while the inefficient delay in Fearon (1998) is mainly due to the dichotomous bargaining choices and (or) uncertainty.

There are fewer theoretical studies of government termination. Laver and Shepsle (1996) stress that the ending of one cabinet begins the formation process for the next and that dissolution and formation are conceptually nonseparable, though their own emphasis is more on the making than breaking of governments. Lupia and Strøm (1995) show that majority governments may dissolve and call early elections when the expected payoff is high enough. Their model explains why a cabinet, which is an "equilibrium" of the earlier bargaining process, might find it worthwhile to terminate its tenure and call an election. All of these studies make important contributions, but fall short of the kind of systematic integration that we see as necessary.

We argue that the lengths of coalition bargaining and government survival are interdependent duration processes. Unfortunately, to this point, the two have been studied largely in isolation. The single equation studies suffer from multiple sources of bias. One potential problem is omitted variable bias in regressions that leave out the important "right-hand-side" duration. Simultaneity is a concern for studies that do connect government formation and dissolution in single equation models by putting variables like crisis duration or the number of cabinet formation attempts on the right-hand-side of government survival regressions. The simultaneity problem is obvious from the structures of these models. Bargaining and survival durations are clearly related, but the causal arrow points both ways. If we put one duration on the right hand side of a model explaining the other-as is frequently done in studies of government survival-our estimates will be biased by the reverse causal relationship. The clear empirical implication of these formal models is that we should not estimate coalition bargaining and government survival durations separately or naively put one duration on the right-hand-side of a single-equation regression that has the other duration on the left-hand-side.

Our dataset consists of 475 cabinets from sixteen Western European countries -Austria, Belgium, Denmark, Finland, France (Fourth Republic), Germany, Iceland, Ireland, Italy, Luxembourg, The Netherlands, Norway and Sweden. The data run between 1945 and 1998.

Table A.1 is the list of variables that are (theoretically) relevant for our purposes. Most of these are common in the existing literature. They are conceptualized and measured in the following ways:

Interestingly, there is a positive relationship between the time it takes for government for-

Table A.1: Variables That Might Affect Each Duration

	Coalition Formation Duration		Government Survival Duration
Country attributes:			
	Investiture		Investiture
*	Continuation		
*	Pre-election coalition		
*	Pre-election coalition×Effective parties		
Party structure attributes:			
	Effective parties		Effective parties
	Polarization		Polarization
	Returnability		Returnability
	Ideological diversity		Ideological diversity
Cabinet attributes:			
	Post-election		Post-election
	Caretaker		Caretaker
	Survival duration		Formation duration
*	Previous defeat	*	Maximum duration
		*	Majority status

* indicates exclusive variables.

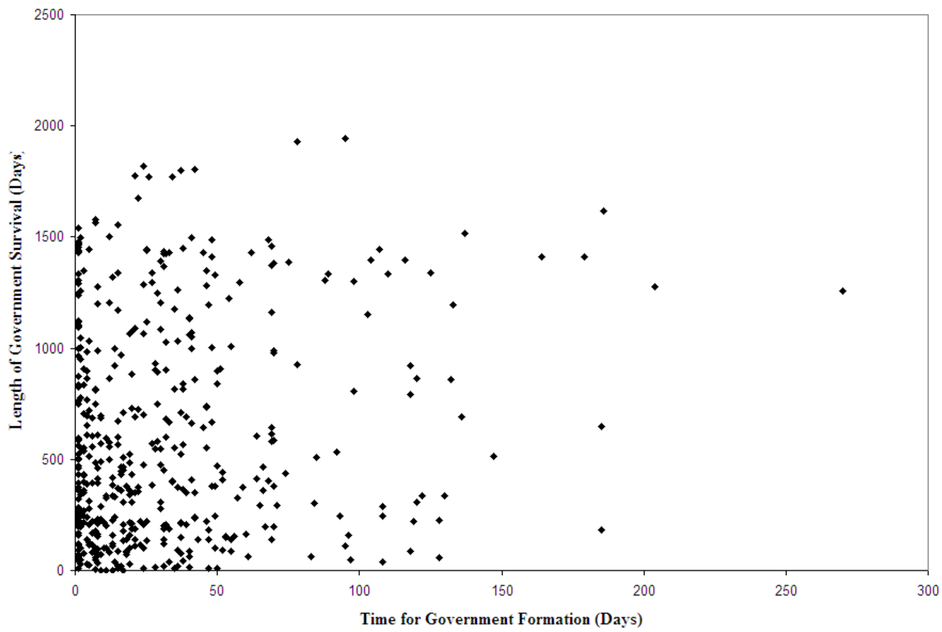
mation and the length of government survival in our sample (see Figure A.1). Governments that formed in less than fifty days survived, on average, 580 days whereas coalitions that took more than 100 days to reach agreement lasted 818 days. This is a bit perplexing since we might expect long delays in government formation to be indicative of the inability of parliamentary parties to work together effectively (King et al. 1990). There is another way to look at this relationship, however. Parties that anticipate long-lasting governments may bargain harder over coalition agreements since these "contracts" will determine the balance of executive power, distribution of benefits from holding office, and overall course of policy for a significant period of time into the future. Is this relationship spurious or causal? And what implications, if any, does this interdependence have for empirical analyses of government formation and survival durations?

The FIML and 2SLS estimators rely on instruments to identify the causal effects of cabinet

Figure A.1: Government Formation and Duration

Time for Formation	Average Survival	Standard Deviation	Min	Max
Less than 50 days	580	481	1	1818
Between 50 and 100 days	649	515	10	1941
More than 100 days	818	529	36	1616

Data source: Warwick (1994), Golder (2005), Keesing's World News Archive



formation duration on government survival and vice versa. We use the continuation and maximum duration variables as instruments. In both cases, we think it highly plausible on theoretical grounds that the instruments satisfy the necessary exclusion restrictions. We present two sets of results for each estimator. The first set is from a covariate sparse specification, and the second is from a covariate rich specification. The sparse specification includes, in addition to the endogenous variables, the instruments needed for the FIML and 2SLS estimators. We focus primarily on the differences between the AEDM and FIML estimators and the simultaneous relationship between cabinet formation duration and government survival.

With the FIML estimator we find robust evidence that the positive correlation between bargaining duration and government survival seems to be driven by the latter causing the former. The covariate-sparse and rich comparison highlights the unbiasedness and efficiency

Table A.2: Estimation Results for the Cabinet Formation and Survival Duration

	Independent Durations		Exogenous Durations		2SLS		FIML	
Formation duration ($\mathbf{y}_1$)								
θ_1 (Scale parameter 1)								
Constant	3.473***	1.877***	2.602***	1.944***	-1.31	0.171	2.356***	1.319***
	(0.063)	(0.249)	(0.157)	(0.281)	(0.998)	(2.604)	(0.265)	(0.416)
Continuation	-0.958***	-0.865***	-0.944***	-0.871***	-1.23***	-0.732***	-1.01***	-0.862***
	(0.14)	(0.135)	(0.134)	(0.136)	(0.214)	(0.175)	(0.138)	(0.135)
Investiture		0.019		0.025		0.266		0.036
		(0.104)		(0.105)		(0.188)		(0.105)
Effective Parties		0.102**		0.1**		0.201***		0.105***
		(0.041)		(0.041)		(0.062)		(0.041)
Polarization		0.882*		0.9*		1.15		1.168**
		(0.462)		(0.463)		(1.068)		(0.494)
Returnability		0.507*		0.505*		0.711*		0.515*
		(0.277)		(0.277)		(0.386)		(0.276)
Post-Election		1.173***		1.215***		1.147***		1.126***
		(0.103)		(0.130)		(0.279)		(0.106)
Caretaker		0.317		0.301		0.241		0.414*
		(0.214)		(0.216)		(0.503)		(0.22)
α_1 Dependency 1								
Survival			1.019***	-0.07	0.704***	0.068	1.036***	0.086*
			(0.083)	(0.136)	(0.172)	(0.383)	(0.088)	(0.051)
λ_1^{-1} (Shape parameter 1)								
Constant	1.187***	1.019***	1.143***	1.019***			1.168***	1.015***
	(0.043)	(0.038)	(0.042)	(0.038)			(0.043)	(0.038)
Cabinet survival ($\mathbf{y}_2$)								
θ_2 (Scale parameter 2)								
Constant	5.207***	6.251***	5.13***	6.271***	5.875***	6.302***	5.302***	6.251***
	(0.106)	(0.182)	(0.108)	(0.197)	(0.420)	(0.264)	(0.118)	(0.182)
Max Duration	0.998***	0.488***	0.719***	0.485***	1.368***	0.473***	0.191***	0.488***
	(0.084)	(0.103)	(0.123)	(0.104)	(0.204)	(0.14)	(0.044)	(0.103)
Investiture		-0.192***		-0.194***		-0.385***		-0.193***
		(0.069)		(0.069)		(0.121)		(0.070)
Effective Parties		-0.026		-0.026		-0.071		-0.026
		(0.024)		(0.024)		(0.053)		(0.025)
Polarization		-1.686***		-1.694***		-2.495***		-1.685***
		(0.25)		(0.252)		(0.501)		(0.25)
Returnability		-0.258		-0.273		-0.312		-0.261
		(0.167)		(0.177)		(0.337)		(0.172)
Post-Election		0.314***		0.315***		0.379		0.312***
		(0.09)		(0.090)		(0.249)		(0.094)
Caretaker		-1.092***		-1.096***		-1.022***		-1.092***
		(0.148)		(0.149)		(0.219)		(0.148)
α_2 Dependency 2								
Formation			0.251***	-0.023	-0.618***	-0.031	-0.053*	0.002
			(0.092)	(0.091)	(0.206)	(0.186)	(0.027)	(0.026)
λ_2^{-1} (Shape parameter 2)								
Constant	0.785***	0.691***	0.781***	0.691***			0.791***	0.691***
	(0.03)	(0.026)	(0.03)	(0.026)			(0.031)	(0.026)
Log-Likelihood	-1499.24	-1369.39	-1479.32	-1369.22			-1491.02	-1367.53

Note: We could compute the Weibull shape parameter λ for the 2SLS model by using the other parameter estimates and the estimated variances. We have not done the computation yet. Significance levels : $*$: 10% $**$: 5% $***$: 1%

advantages of the FIML in small samples. The AEDM estimator finds a positive and statistically significant relationship between cabinet duration formation and government survival in both equations. This is not surprising given the positive covariance, but keep in mind that the *causal* argument that prolonged formation processes (what many scholars refer to as the "crisis" duration) lead to longer-lived governments is viewed by most as dubious. We expect the opposite relationship, and this is what we find with the FIML estimator. Turning to the covariate rich specification, the relationships between the government formation and survival durations is wiped away in the AEDM estimates. By contrast, with the FIML estimates, we continue to find a statistically significant and positive effect of government survival on formation duration. This is due in large part to the efficiency of the estimator. The estimated standard error for the FIML coefficient on the government survival variable is almost one-third the size of the AEDM standard error.

Overall, we interpret the FIML results as strong evidence that parties anticipate the length of the future government's tenure and this affects how they bargain. This is the idea of strategic interdependence that comes out of the game theoretic literature on the topic developed by Diermeier, Merlo and others. We do not find evidence of the reverse causal relationship. In other words, although there are studies that suggest that the duration of formation processes affects government survival, we do not find robust evidence that this is the case. These theories maintain that longer bargaining indicates the difficulty in reaching agreements among the coalition members in general and hence portends a shorter lifespan for the formed government.

Duration Interdependence Across Actors: Strategic Timing of Issue Position Taking

It is often said that timing is everything in politics. This is certainly true when it comes to the behavior of elected politicians, and position taking on legislation is one of the clearest examples. Drawing on the logic of formal signaling models, Box-Steffensmeier, Arnold and Zorn (1997) argue in their seminal paper that issue position taking in Congress will be strategically timed. Members of Congress (MCs) who receive clear signals about the policy preferences of their constituents will announce early, while those who receive mixed signals will delay. They also contend that constituency preferences will interact with individual-level factors in either cross-cutting or reinforcing ways, and that institutional factors such as leadership status and committee membership will influence the timing of issue position taking. In their empirical analysis, Box-Steffensmeier et al. examine issue position taking on NAFTA. They find that MCs from border districts took early positions, as did Republican leaders, ceteris paribus. Conservative MCs from highly unionized districts delayed their position taking.

Boehmke (2006) extends this analysis by linking the timing of issue position taking by MCs and the content of their positions through unobservables. He argues that factors that cause delay in position taking on NAFTA also make it more likely that members will support the legislation. Legislators that hold out, whatever the reason may be, ultimately decide to vote in favor, either because of presidential pressure or out of concern for party interests. Darmofal (2009) develops the model further by allowing for spatial correlation in the timing of issue position taking among representatives. He models spatially connected individual and shared (state-level) frailties. In his preferred specification, the state-level shared frailties specification, MCs from the same state have a common random effect, and these effects are geographically correlated. Frailties cluster among states that share borders.

The strategic nature of issue position taking is clear from this literature. One form of interdependence that is left out of these models is strategic interdependence across members of Congress. Darmofal's models come the closest to capturing this interdependence, but his model is better interpreted as capturing omitted variables that cluster geographically rather than true interdependence among MCs since the correlation is only in the disturbance term. There is good reason to expect strategic interdependence across members, particularly members from districts in close proximity. These members represent overlapping constituencies, and therefore may have incentives to take early positions to signal their commitment and resolve on a particular issue. If true, this would spark competitive dynamics among representatives. It is also possible that members of Congress free-ride off of the early position taking of their colleagues. Early position taking and the political responses it provokes provide valuable information to other members who, at some point, will be expected to take a stance. These relationships can be modeled using our interdependent durations model.

The dependent variable is the number of days after August 11, 1992, the date when the first

MC (Peter Visclosky) announced his NAFTA position, before the other MCs took a pro or con position on the legislation. The constituency variables included in the analysis are the district-level Perot vote, union membership, and average household income. The interest group factors include the contributions from corporate and labor PACs. The institutional variables are NAFTA committee membership and party leadership indicators. Ideology is the individual-level variable, which is interacted with constituency-level variables to model cross-cutting pressures on MCs. Interestingly, there is a substantial amount of within state variation in the data. In the case of logged durations, the within variance is larger than the between, or, more specifically, the average within state variance is larger than the variance in state-level means. This fact is consistent with negative interdependence or free-riding behavior.

Since we are analyzing a single duration, our interdependent durations model simplifies to a spatial-lag model. We use Darmofal's adjacency matrix based on queen contiguity to connect MCs and include state-level fixed effects for all states with more than one representative. The same four estimators used previously–the assumed independent, assumed exogenous, 2SLS, and FIML estimators–can be applied to this model. We report the results in Table A.3. The first column gives the Box-Steffensmeier et al. results from their Cox proportional hazards model. The remaining four columns give the estimates for our Weibull accelerated failure time models.[2] In all of the models that allow for duration interdependence, we find evidence of free-riding (i.e., the coefficient estimate for rho is always negative). When one MC for whatever reason-including constituency, institutional, and individual-level factors-announces an early position on NAFTA, his or her colleagues from bordering districts are more likely to delay their position taking. As expected, the 2SLS estimator has the largest standard errors, and the assumed exogenous estimator overstates the strength of interdependence relative to

[2]Note that, to be consistent, the coefficient estimates from the proportional hazards and accelerated failure time (AFT) models should have the opposite signs. The proportional hazards model gives the effects of covariates on the hazard rate, while the AFT model gives the effects of covariates on the expected time until failure.

the FIML.

Table A.3: The Timing of NAFTA Position Taking

	BSAZ (1997)	Independent	Exogenous	2SLS	FIML
Constituency Factors					
Union Membership	3.21**	-0.33	-0.34	1.72**	-0.33
	(1.19)	(0.28)	(0.28)	(0.87)	(0.28)
Perot Vote, %	-4.91	0.34	0.39	1.69	0.39
	(4.27)	(0.57)	(0.56)	(1.88)	(0.56)
Perot Vote, % Squared	15.64	-1.18	-1.48	-5.36	-1.44
	(11.72)	(1.55)	(1.55)	(5.32)	(1.55)
Mexican Border	1.84**	-0.24***	-0.27***	-0.64***	-0.27***
	(0.32)	(0.05)	(0.05)	(0.15)	(0.05)
Household Income	0.01	0.01	0.01	0.02	0.01
	(0.09)	(0.01)	(0.01)	(0.04)	(0.01)
Interest Group Factors					
Corporate Contributions	-1.44**	0.10	0.09	0.25	0.09
	(0.52)	(0.07)	(0.07)	(0.22)	(0.07)
Labor Contributions	1.09*	-0.09	-0.10	-0.02	-0.09
	(0.50)	(0.06)	(0.06)	(0.20)	(0.06)
Institutional Factors					
NAFTA Committee	0.04	-0.0004	0.002	0.02	0.002
	(0.11)	(0.0130)	(0.013)	(0.04)	(0.013)
Republican Leadership	0.56**	-0.06*	-0.05	-0.05	-0.05*
	(0.26)	(0.03)	(0.03)	(0.10)	(0.03)
Democratic Leadership	0.08	-0.02	-0.02	-0.03	-0.02
	(0.23)	(0.03)	(0.03)	(0.09)	(0.03)
Individual Factors					
Interaction Effect of Ideology and	-4.39**	0.44**	0.42*	0.61	0.42*
Union Membership	(1.78)	(0.22)	(0.22)	(0.73)	(0.22)
Interaction Effect of Ideology and	0.16	-0.02	-0.01	-0.006	-0.01
Household Income	(0.13)	(0.01)	(0.01)	(0.049)	(0.01)
Shape Parameter λ		8.92***	8.95***		8.95***
		(0.38)	(0.38)		(0.38)
Spatial Parameter ρ			-0.09**	-0.69**	-0.08*
			(0.05)	(0.32)	(0.04)
Log Likelihood		194.277	196.300		196.041

This table compares our results with Table 2 in Box-Steffensmeier et al. (1997). Table 2 in Box-Steffensmeier et al. (1997) is for the model that explains the timing of position taking by the House members. "Stars" represent the following; $*$: 10% $**$: 5% $***$: 1%. All of our models were estimated with state fix effects.

A.2 Appendices of Essay 2

A.2.1 Stata Code to Estimate Simultaneous Duration Model Accounting for Right-Censoring

```
**********************************************************************
```

```
* Program to estimate SDEQ (Weibull) Model Accounting for Right-Censoring
*************************************************************************
* Last updated July 31, 2012, by Aya Kachi
* Stata ver.9
***********************************************************
clear
pr drop _all
set more off

***********************************************************
* Likelihood evaluator
*----------------------------------------------------------
* Model with two dependent durations
***********************************************************
program define sdeq_cvt_copula_hybrid_ll
args lnf mu1 mu2 alpha1 alpha2 lambda1 lambda2
tempvar J logJ alpha ay1 ay2
scalar a1 = `alpha1'
scalar a2 = `alpha2'
scalar l1 = `lambda1'
scalar l2 = `lambda2'
matrix I = [1, 0 \ 0, 1]
matrix A = [0, a1 \ a2, 0]
matrix IA = I-A
matrix L = [l1, 0 \ 0, l2]
matrix IAL = IA*L

gen `ay2' = `alpha1'*$ML_y2
gen `ay1' = `alpha2'*$ML_y1

qui replace `lnf'= (1-censored)*(`lambda1'*($ML_y1-`ay2'-`mu1') - /*
*/ exp(`lambda1'*($ML_y1-`ay2'-`mu1')) +  `lambda2'*($ML_y2-`ay1'-`mu2') /*
*/ - exp(`lambda2'*($ML_y2-`ay1'-`mu2')) + ln(abs(det(IAL)))) /*
*/ + censored*(ln(1-`lambda1'*(1 - exp(-exp(`lambda1'*($ML_y1-`ay2'-`mu1')))) - /*
*/ `lambda2'*(1 - exp(-exp(`lambda2'*($ML_y2-`ay1'-`mu2'))))  /*
*/ + (1 - exp(-exp(`lambda1'*($ML_y1-`ay2'-`mu1'))))*(1 - /*
*/ exp(-exp(`lambda2'*($ML_y2-`ay1'-`mu2'))))*abs(det(IAL))))

end

***********************************************************
* Load data
***********************************************************
drop _all
use
"PATH TO DATA", clear

* for analysis without mauritius
drop if survival == 34

gen gdppc_sq_lib = rgdp96pc_gled_100000_lib^2
gen gdppc_sq_rev = rgdp96pc_gled_100000_rev^2

gen lnform = ln(transition)
```

```
gen lndur = ln(survival)

global Y1 lnform
global Y2 lndur
global X1 rgdp96pc_gled_100000_lib gdppc_sq_lib commonwealth_2007_lib /*
  */ urbanpercent_10percent_lib moslem_pacl_10percent_lib dic_military fuelexp_10percent_lib
global X2 rgdp96pc_gled_100000_rev commonwealth_2007_rev urbanpercent_10percent_rev /*
  */ moslem_pacl_10percent_rev dic_military dic_noindep president

***********************************************************
* Produce starting values by a univariate Weibull duration
* model for each duration equation
***********************************************************
stset transition
streg $X1, dist(weibull) time
matrix stregbp1 = e(b)
local col1 = colsof(stregbp1)
matrix stregb1 = stregbp1[1,1..`col1'-1]
matrix coleq stregb1 = mu1
local stregp1 = 1/exp(stregbp1[1,`col1'])
stset survival
streg $X2, dist(weibull) time
matrix stregbp2 = e(b)
local col2 = colsof(stregbp2)
matrix stregb2 = stregbp2[1,1..`col2'-1]
matrix coleq stregb2 = mu2
local stregp2 = 1/exp(stregbp2[1,`col2'])

***********************************************************
* Estimate SDEQ (Weibull) model
***********************************************************
ml model lf sdeq_cvt_copula_hybrid_ll (mu1: $Y1=$X1) (mu2: $Y2=$X2) (alpha1:) (alpha2:) /*
  */(lambda1:) (lambda2:)
ml init stregb1
ml init stregb2
ml init alpha1:_cons=1
ml init alpha2:_cons=-1
ml init lambda1:_cons=`stregp1'
ml init lambda2:_cons=`stregp2'
ml max

* to compute aic and bic
gen double aic = -2*e(ll)+2*e(rank)
gen double bic = -2*e(ll) + log(32) * e(rank)
```

A.2.2 Stata Code to Estimate the Univariate Weibull Duration Models with Right-Censoring

```
*Program to Estimate Univariate Weibull Duration Model (Right-Censoring)
clear
pr drop _all
set more off
```

```
**************************
*Open Data for Regression
**************************
drop _all
use "PATH TO DATA", clear

drop if survival == 34
gen gdppc_sq_lib = rgdp96pc_gled_100000_lib^2
gen gdppc_sq_rev = rgdp96pc_gled_100000_rev^2

gen lnform = ln(transition)
gen lndur = ln(survival)
global Y1 lnform
global Y2 lndur
global X1 rgdp96pc_gled_100000_lib gdppc_sq_lib commonwealth_2007_lib /*
*/ urbanpercent_10percent_lib moslem_pacl_10percent_lib  dic_military /*
*/ fuelexp_10percent_lib survival
global X2 rgdp96pc_gled_100000_rev commonwealth_2007_rev /*
*/ urbanpercent_10percent_rev moslem_pacl_10percent_rev dic_military /*
*/ president dic_noindep transition

************************
*Produce starting values
************************
stset transition
streg $X1, dist(weibull) time

gen double aic1 = -2*e(ll)+2*e(rank)
gen double bic1 = -2*e(ll) + log(32) * e(rank)

stset survival, failure(censored==0)
streg $X2, dist(weibull) time

gen double aic2 = -2*e(ll)+2*e(rank)
gen double bic2 = -2*e(ll) + log(32) * e(rank)
```

A.3 Appendices for Essay 3

A.3.1 Suggested Uses of Social Network Analysis (SNA) for Estimated Implicit Networks Among Countries

Estimated Latent Regime Interdependence among Countries

With spatial econometric models, it is usually difficult to grasp the true meaning of effects simply by looking at estimated coefficients, such as those reported in Table 4.3. For example, as mentioned in preceding sections, the regime interdependence can be captured only by a product of the estimated coefficient and the given spatial weights, $\hat{\rho}_r \mathbf{W}_r$ or $\hat{\gamma}\mathbf{L}$, not simply by the coefficients $\hat{\rho}_r$ and $\hat{\gamma}_r$. A weights matrix $\mathbf{W}_r$ indicates how strongly a pair of countries is interconnected and the coefficient $\hat{\rho}_r$ tells us how much this particular connectivity matters in explaining countries' democracy scores. It is sometimes useful to compute the overall interdependence of countries by $\sum_{r=1}^{R} \hat{\rho}_r \mathbf{W}_r + \hat{\gamma}\mathbf{L}$. This quantity is the estimated latent structure of interdependence that explains and is partially explained by the level of democracy of the included countries.

Most works in spatial studies report the $\sum_{r=1}^{R} \hat{\rho}_r \mathbf{W}_r$ matrix to demonstrate the estimated interdependence, but do not study the characteristics of the information contained in the connectivity matrix. (Hays, Kachi and Franzese (2010) is an exception.) This practice is somewhat ironic given that spatial studies are motivated by the very notion that there are a number of political outcomes and behavior that cannot be explained only by the actors' attributes; the strength of ties among actors is also an important determinant of such phenomena. The following graphs exemplify some of the patterns of the implicit network interdependence in democracy scores that were revealed by the estimation.

Most countries have at least weak connections with some countries, which makes the network graphs tend to look very busy. Figure A.2 demonstrates only the ties with high magnitude (the 95th-percentile) among all the statistically significant ties at the 99% confidence level. Figure A.2-(a) (the top panel) shows such ties in the time period of 1971-1975 and (b) (the bottom panel) is for the 1996-2000 time period. The size of edges indicates the strength of the estimated ties. Since the estimated coefficient of the similarity matrix $\mathbf{L}$ is a negative value (-0.104), some cells of the overall interdependence $\sum_{r=1}^{R} \hat{\rho}_r \mathbf{W}_r + \hat{\gamma}\mathbf{L}$ have negative values; however it turns out that all the negative dependencies are statistically not significant, leaving in only the positive interdependence in the plots. The node size represents the democracy level (larger nodes for higher democracy scores), and the node color indicates geographical region.[3] For example, an arrow from Japan to the Philippines indicates the influence of Japan on the Philippines, or the Philippines "learning from Japan's regime experience".

As an example of over-time network change, Figure A.3 shows the ego network of Thailand for each of the nine time periods. The ego network is defined as a network that consists of a focal actor ("ego" and in this case it's Thailand) and the actors to which ego is directly connected, as well as the ties among these "alters". A big (red) circle in each plot indicates the location of Thailand. There are various ways to analyze these plots. One might be interested in the change in network density over time. Since the number of countries to which Thailand is connected generally increases over time, the plots become "messier"; however, we never know if the density of networks is going up or down over time until we compute it for each year. Network density is commonly measured by the number of ties divided by the number of possible pairs–in other words it is a measure of the proportion of ties formed out of all the possible combinations in a given network. For the case of Thailand, the density has been almost constantly dropping since the 70's (see Table A.4), even though the graphs might look increasingly busy over time.

[3] Black: Western Europe. Blue: Central and South America. Green: Middle East. Light blue: Oceania. Light green: Asia. Gray: Eastern and Central Europe. Red: North America. Pink: Africa.

Figure A.2: Estimated Latent Regime Dependencies: High-Magnitude (95 Percentile) Cases for the 1971-1975 and the 1996-2000 Time Periods

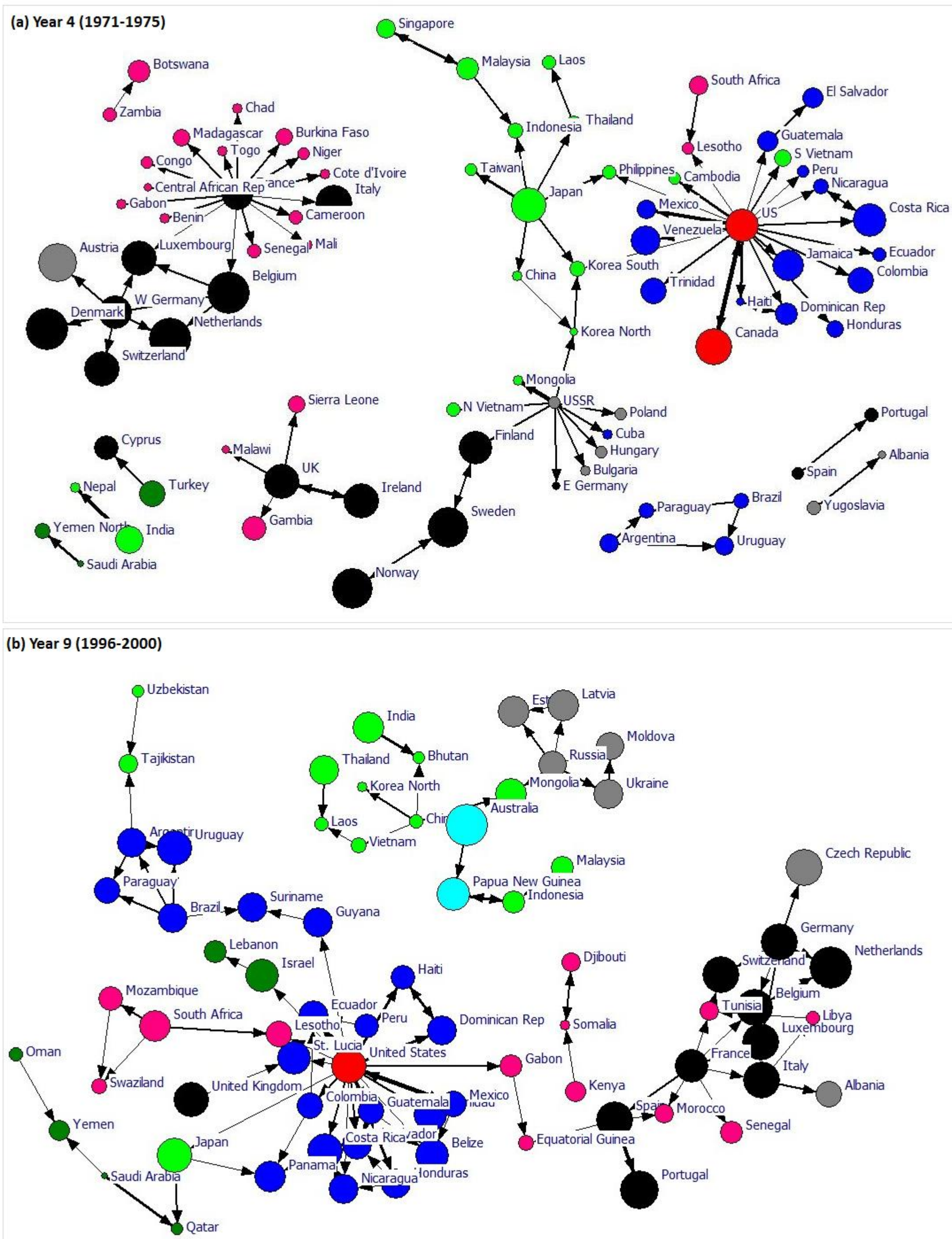


Figure A.3: Thailand's Ego Networks (Distance = 1)

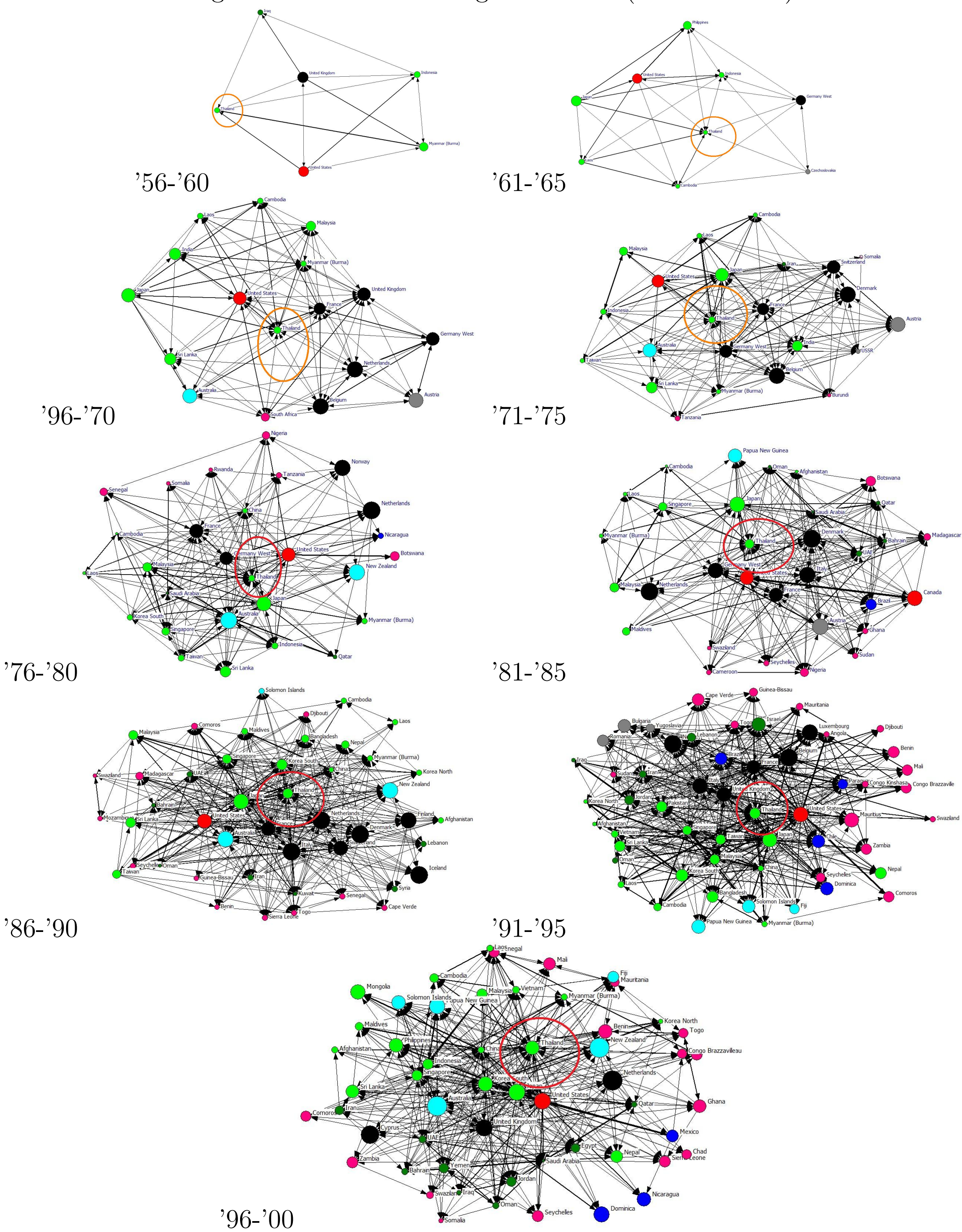


Table A.4: Density of Thailand's Ego Networks

	'56-60	'61-65	'66-70	'71-75	'76-80	'81-85	'86-90	'91-95	'96-00
Density	50.00	41.07	49.17	41.34	28.92	30.65	22.68	20.61	15.58

Note: Densities are computed using UCINET ver.6.

A.3.2 Research Note on a New Study of Path Dependency in Democratization

Motivation and Research Question

The main purpose of Essay 3 was to show empirically that countries do form invisible diplomacy ties among them based on their regime similarity (*selection*), and countries that are connected through this network support or approve of each other's regimes (*contagion* in a passive manner). Over time, regime-support networks across countries and the political regimes of these countries coevolve, and regimes are reinforced over time. Furthermore, the simulation analysis in Section 4.11 indicates that only with the regime-reinforcement mechanism, countries' democracy/autocracy levels are sensitive to the initial set of democracy levels of all the countries in the world.

However, this initial condition sensitivity is only "suspected" from a set of simulation analyses: there were no explicit statistical tests for the initial condition sensitivity in the essay. In the next iteration of my democratization studies, I hypothesize that the self-selecting (i.e., regime-approving) network formation among countries and regime reinforcement through these networks make the system of political regimes *path dependent*, and provide the empirical evidence that supports this hypothesis.

In the paper, I first provide a rigorous, mathematical concept of *path dependency*, based on the definition in Page (2006). In fact, by this definition, *path dependency* is the most strict form of the three types of *history dependency*. In the literature of historical institutionalism, the term path dependency is used in various ways. All three types of history dependencies are called “path dependency” in the literature. The substantive theory of democracy levels presented in this paper is the same as the one presented in Essay 3. In order to test for the hypothesis of path dependence specifically, I introduce a new statistical model, a spatial-and-networks hybrid model based on Franzese, Hays and Kachi (2012). In short, this particular model can test for the existence of path dependency, because the model specification (Markov state-space models with the logit probability functions) is directly based off of a mathematical model that describes the connection between the reinforcing mechanism and the existence of path dependency, or multiple equilibria. This formal theoretical model provides us with the parameter conditions under which path dependence exists.

What are Path Dependency and History Dependency?

The term “path dependence” has been used in political science without a concrete concept and definition on which scholars can agree. Partly for this reason, the literature is hard to understand. It is not clear whether one disagrees with others because of the definition of the concept or insufficient evidence for a given definition of “path dependence.” In this study, I subscribe the mathematical definition of history dependency summarized by Page (2006). According to Page (2006), *history dependence* broadly refers to phenomena in which past conditions changes the course of a system in the future. This broad notion is often mistaken and referred to as “path dependency” in the literature. Page (2006), however, defines history dependence most broadly and differentiate three increasingly restrictive subcategories of history dependence: *state*, *phat*, and *path dependence*. The most-restrictive *path dependence* means that a system’s future history depends on the *path*, i.e., the *sequence* or *order*, of past conditions, and not merely on the set of past events. When a system is sensitive to a set of events that occurred in the past but not the order of those events, it is a less-restrictive form of history dependence, *phat dependence*. The least-restrictive *state* dependence is where a system’s trajectories can be partitioned into a finite number of states that contain all relevant information for the future of the system regardless of events outside that partition (meaning that the system's future depends on its current state, not the path or set of earlier conditions).

From now on, I am going to use this minimalist but mathematically rigorous definition of history dependence, which includes a stricter concept of path dependence. One of the merits of these definitions is that they are applicable for both quantitative and qualitative analyses of history dependence regardless of substantive contexts. They seem appropriate also because the recent mini symposium on path dependency published in *Political Anlaysis* attempts to set the common ground both for theoretical and empirical analysts of path dependence, using Page (2006)'s definition.[4]

"Path Dependency" in Political Science

In political science, there have not been rigorous quantitative empirical studies of path dependency, or even of history dependence (whatever researchers mean by their use of "path dependency"), for any substantive topics.[5] Most of existing studies that touch upon issues related to path dependence are qualitative, coming from the historical institutionalism literature (e.g., Collier and Collier 1991; Mahoney 2000, 2001; Pierson 2000; Pierson and Skocpol 2002; Pierson 2004; Thelen 1999; Waisman 1987). The research method—qualitative vs. quantitative—aside, The primary problem in summarizing and advancing the literature is the fact that the definition of "path dependency" has been arbitrary and varying from study to study. The concept of path dependence, among existing works, that is most rigorous and closest to what I use in this study is the one presented by Pierson and Skocpol (2002). They explain that "outcomes at a 'critical juncture' trigger feedback mechanisms that reinforce the recurrence of a particular pattern into the future." Yet this explanation also lacks the

[4]For details of the symposium, see *Political Anlaysis* (Vol.20(2)) featuring Freeman and Jackson (2012), Bedner, Page and Toole (2012), Jackson and Kollman (2012) and Franzese, Hays and Kachi (2012).

[5]This is only a preliminary and very minimal set of literature reviews to prove the main point that the definition of path dependence is unclear in existing studies, and there have not been any rigorous quantitative tests for path dependent political institutions under any definitions.

distinction between the effects of simply earlier outcomes/events (i.e., the possibility of *phat* dependency) and the effect of the order of past events in the history (i.e., the possibility of the true path dependency).

The literature is even thinner once we focus on the topic of democratization. Alexander (2001) presents an article titled "Institutions, Path Dependence, and Democratic Consolidation", for example; however, the notion of path dependency in this paper is defined by whether a given (democratic) institution is "locked in" or not in the long run. The author shows, by narratives, that not all democratic institutions are locked in—formal institutions are less likely to be locked in during the course of their development. Hence, he claims, democratic consolidation is not necessarily path dependent. How can we learn that an institution is "locked in" in a way historical institutionalists conceptualize? There is no statistical test for it.

Another noteworthy study is Mainwaring and Pérez-Liñán (2008). The theme of this study is more similar to what I think as a potential history dependency of democracies. Focusing on Latin American countries, the authors point out the variation in the current democracy levels in the region and ask what contributed to the varying democracy levels of all these countries, which democratized during the "third wave" of democratization between 1978 and 1992. They hypothesize that the country's "past history of political regimes"—what they call regime heritage or legacy—is the key determinant. They explicitly claim that this study is different from the typical "path dependency" studies—this is a "regime legacy" study. Again, the authors seem to define path dependence as a process only relevant to the course of system development, in which institutions make "major" or "dramatic" shift over time. They claim that they are interested in more subtle and flexible notion of institutional changes in explaining democratization. This study is one of the few quantitative empirical studies that test for the effects of earlier democracy levels on the countries' later democracy levels. To me, this is a study of history dependence, because the mechanism for which they attempt to test is a simpler type of history dependence—*phat* dependence.

However, Mainwaring and Pérez-Liñán (2008)'s theory has a major shortcoming that the authors also admit in the article. Their study contributes to test whether past democratic experiences positively affect the countries' democracy levels in the future, but they do not specify or test for the mechanism that links the earlier regime conditions and the current level of democracies. The authors point out that both their "regime legacy" study and existing "path dependency" studies of political institutions share this weakness. They state;

> Thelen (1999) argued that one shortcoming of some work that invokes the notion of path dependence is that it fails to explain the mechanisms that generate such dependence. A regime heritage argument faces the same challenge; it must explain why regime legacies shape the current level of democracy. The answer is not immediately obvious.

Contributions of My Study

The contribution of my study suggested here is threefold. First, I will introduce a mathematically rigorous definition of path dependence based on Page (2006). This will enable researchers in any subfields to discuss the process by the mathematical characteristics of the process and not by "how the process looks to each analyst." Second, my study of democracy and path dependence is free of the lack-of-mechanism problem in the existing regime legacy and "path dependent" studies of political institutions. In fact, my theory starts from a possible *mechanism* that can general history dependence, including the rigorous notion of path dependence. The mechanism presented and tested here is regime reinforcement across countries and over time. Third, I suggest a statistical methodology to test for the existence and the strength of thes regime-reinforcing mechanism, enabling us to conclude whether democratization is a path dependent process of not, at least in terms of this particular mechanism.

Empirical Strategy: A Spatial-Lag-and-p^* Hybrid Model

I introduce the spatial-and-network hybrid model suggested in Franzese, Hays and Kachi (2012). This model consists of two relatively simple logistic discrete-time Markov models. I explain each part of the empirical model, by matching it with each component of my theoretical model for democracy levels. Note that, in this model, both variables for the regime and the tie-formation decision are dichotomous. Country i is either democracy or autocracy in a given time period ($s_{i,t} = 1$ is democracy and $s_{i,t} = 0$ is autocracy), and each pair of countries (countries i and j) either forms or does not form a regime-support tie in a given time period t ($d_{ij,t} = 1$ means forming a tie).

$$\begin{cases} Pr(s_{i,t} = 1|\mathbf{s}_{t-1}, \mathbf{d}_{t-1}) = logit(\beta_0 + \beta_1 s_{i,t-1} + \beta_2 \mathbf{d}_{i,t-1}\mathbf{s}_{t-1}) \\ Pr(d_{ij,t} = 1|\mathbf{s}_{t-1}, \mathbf{d}_{t-1}) = logit(\gamma_0 + \gamma_1 d_{ij,t-1} + \gamma_2 \cdot I(s_{i,t-1} = s_{j,t-1})), \end{cases} \tag{A.15}$$

The first equation predicts the probability that a country takes a certain "behavior"—democracy, conditioned on all the other countries regimes types in the past time period and all the pairs' tie-forming decisions in the past time period. First, this probability is explained (see the right-hand side) by β_0, the average tendency for country i to be democratic. This is a country fixed effect. It also depends on $\beta_1 s_{i,t-1}$ the country's own political regime in the previous time period. If it was a democracy in the previous time period, it is more likely to be democratic than autocratic in the following period. The last term $\beta_2 \mathbf{d}_{i,t-1}\mathbf{s}_{t-1}$ is one represents the contagion of political regimes. The component $\mathbf{d}_{i,t-1}$ is a row vector of size N that includes the $(N - 1)$ pairwise tie-formation indicators between i and each of all the other states at the end of period $t - 1$. This is multiplied by the regime type (1: democracy and 0: autocracy) of the other corresponding states'. This term measures the strength of the effect of the weighted average of all the other democracy on country i's tendency to be a democracy. This equation is, in fact, a simple spatial-lag logit model.

Next, the second equation predicts the probability that a dyad of two countries i and j form a dependency tie. This probability is explained first by γ_0, the average baseline tendency for this pair to form a tie. It also depends on $\gamma_1 d_{ij,t-1}$, the dyad's past tie formation decision. γ assesses how easy it is for a pair to maintain a tie when the tie existed before. The highlight of the model is the very last term. Based on the selection (regime-support) theory, this terms states that if i and j's regimes were the same in the previous time period, then it is more likely that the given dyad form a support tie, and γ_2 captures to what extent this selection dynamic exists. If you are familiar with statistical tools in social network analysis, then this equation is, in fact, the simplest p^* logit model.[6]

Since both the spatial-lag logit and the p^* logit moodel are well defined, estimating (A.15) is straightforward. The two processes, the regime maintaining probability and tie-formation, can be estimated separately or as a seemingly unrelated system of logit equations. If the disturbances are correlated across equations, this approach produces consistent estimates of parameter values, though inefficient. Standard error estimates would be inaccurate. One can correct the standard error estimates using a systems "sandwich" estimator of the variance-covariance matrix.[7]

A Note on ERGM-Based *SIENA* vs. the Spatial-and-Network Hybrid Model

For those who are familiar with the SNA approach, the model specification in the previous section might also look familiar. It is similar to the only other statistical model for actor-oriented coevolution models developed by Snijders and his colleagues (Snijders 2005; Snijders, Steglich and Schweinberger 2007). This model can be estimated by a software also developed by them, called *SIENA* (Steglich, Snijders and West 2006). Their model also intends to distinguish statistically the selection mechanism from the traditional contagion

[6] A p^* model is a random graph model that explains the probability of a dyad's or a unit's forming a network tie.

[7] The sandwich matrix in this formulation, the outer product of the gradients of the likelihoods, provides estimates of the parameter covariances across equations, which are incorporated into the variance estimates.

mechanism. The main difference between their approach and the spatial-p^* approach is in the estimation strategy. Snijders et al.'s actor-oriented coevolution model is based on an Exponential Random Graph Model (ERGM). The statistical inference of ERGM asks what are the chances that the exact observed network structure emerged by chance among all the possible permutations of network structures given the nodes, as opposed to by the node and edge characteristics described "on the right-hand side" of the equation. The number of possible network structures grows faster than exponentially by an increase in the number of included nodes. Estimation of ERGM based models, including Snijders et al.'s actor-oriented coevolution model, is done by simulation. The precision tests are done by permutation, rather than by an analytical computation based on distributional assumptions.

References

Acemoglu, Daron and James A. Robinson. 2006*a*. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Acemoglu, Daron and James A. Robinson. 2006*b*. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Agency, Central Intelligence. Date accessed: July 15, 2012. "The World Factbook." `https://www.cia.gov/library/publications/the-world-factbook/index.html`.

Alemán, José. and David D. Yang. 2011. "A Duration Analysis of Democratic Transitions and Authoritarian Backslides." *Comparative Political Studies* 44(9):1123–1151.

Alexander, Gerard. 2001. "Institutions, Path Dependence, and Democratic Consolidation." *Journal of Theoretical Politics* 13(3):249–270.

Almond, G.A. and S. Verba. 1965. *The civic culture*. Little, Brown.

Alt, James E. and Gary King. 1994. "Transfers of Governmental Power: The Meaning of Time Dependence." *Comparative Political Studies* 27(2):190–210.

Arat, Z.F. 2003. *Democracy and human rights in developing countries*.

Beck, Nathaniel and Johnathan N. Katz. 1995. "What to do (and not to do) with time-series cross-section data." *American Political Science Review* pp. 634–647.

Beck, Nathaniel and Johnathan N. Katz. 1996. "Nuisance vs. substance: Specifying and estimating time-series-cross-section models." *Political analysis* 6(1):1–36.

Beck, Nathaniel, Johnathan N. Katz and Richard Tucker. 1998. "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable." *American Journal of Political Science* 42(4):1260–1288.

Beck, Nathaniel, Kristian S. Gleditsch and Kyle Beardsley. 2006. "Space is more than geography: Using spatial econometrics in the study of political economy." *International Studies Quarterly* 50(1):27–44.

Bedner, Jenna, Scott E. Page and Jameson L. Toole. 2012. "Revised-Path Dependence." *Political Analysis* 20(2):146–156.

Bennett, D. Scott. 1999. "Parametric Models, Duration Dependence, and Time-Varying Data Revisited." *American Journal of Political Science* 43(1):256–270.

Boehmke, Frederick J. 2006. "The Influence of Unobservable Factors on Position Timing and Content in the NAFTA Vote." *Political Analysis* 14(4):421–438.

Boehmke, Frederick J., Daniel S. Morey and Megan L. Shannon. 2006. "Selection Bias and Continuous-Time Duration Models: Consequences and a Proposed Solution." *American Journal of Political Science* 50(1):192–207.

Boix, C. and S.C. Stokes. 2003. "Endogenous democratization." *World Politics* pp. 517–549.

Boix, Carles. 2003. *Democracy and Redistribution*. Cambridge University Press.

Boix, Carles and Sebastian Rosato. 2001. *A Complete Dataset of Political Regmes, 1800–1999*.

Bollen, K.A. 2001. "Cross-National Indicators of Liberal Democracy, 1950-1990." *Codebook. University of North Carolina, Chapel Hill, NC* .

Bowman, K., F. Lehoucq and J. Mahoney. 2005. "Measuring political democracy: Case expertise, data adequacy, and Central America." *Comparative Political Studies* 38(8):939–70.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press.

Box-Steffensmeier, Janet M., Laura W. Arnold and Christopher J.W. Zorn. 1997. "The Strategic Timing of Position Taking in Congress: A Study of the North American Free Trade Agreement." *American Political Science Review* 91(2):324–338.

Bratton, Michael and Nicolas van de Walle. 1992. "Popular Protest and Political Reform in Africa." *Comparative Politics* July:419–442.

Brinks, D. and M. Coppedge. 2006. "Diffusion is no illusion: Neighbor emulation in the third wave of democracy." *Comparative Political Studies* 39(4):463.

Brown, Stephen. 2005. "Foreign Aid and Democracy Promotion: Lessons from Africa." *The European Journal of Development Research* 17:179–198.

Carter, David B. and Curtis S. Signorino. 2010. "Back to the future: Modeling time dependence in binary data." *Political Analysis* 18(3):271–292.

Chapman, Thomas and Philip G. Roeder. 2007. "Partition as a Solution to Wars of Nationalism: The Importance of Institutions." *American Political Science Review* 101(4):677–691.

Cheibub, José Antonio. 2007. *Presidentialism, Parliamentarism, and Democracy*. Cambridge University Press.

Coleman, J.S. 1958. "Relational Analysis: The Study of Social Organizations with Survey Methods." *Human Organization* 17(4):28–36.

Collier, Ruth Berins and David Collier. 1991. *Shaping the Political Arena: Critical junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton University Press.

Coppedge, M. and Reinicke W.H. 1991. *Measuring Polyarchy*. Transaction Pub.

Darmofal, David. 2009. "Bayesian Spatial Survival Models for Political Event Processes." *American Journal of Political Science* 53(1):241–257.

Diermeier, Daniel, Hülya. Eraslan and Antonio Merlo. 2003. "A Structural Model of Government Formation." *Econometrica* pp. 27–70.

Diermeier, Daniel and Peter van Roozendaal. 1998. "The Duration of Cabinet Formation Process in Western Multi-Party Democracies." *British Journal of Political Science* 28:609–626.

Diermeier, Daniel and Randy T. Stevenson. 1999. "Cabinet Survival and Competing Risks." *American Journal of Political Science* pp. 1051–1068.

Dietrich, Simone and Joseph Wright. 2011. "Foreign Aid and Political Development in Africa." *Working paper* .

Epstein, D.L., R. Bates, J. Goldstone, I. Kristensen and S. O'Halloran. 2006. "Democratic transitions." *American Journal of Political Science* pp. 551–569.

Fearon, James. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52(2):269–306.

Fortna, Virginia. 2004. "Interstate Peacekeeping: Causal Mechanism and Empirical Effects." *World Politics* 56:481–519.

Fowler, James H., Michael T. Heaney, David W. Nickerson, John F. Padgett and Sinclair Betsy. 2011. "Causality in Political Networks." *American Political Research* 39(2):437–480.

Franzese, Robert J. and Jude C. Hays. 2006. "Strategic interaction among EU governments in active labor market policy-making: Subsidiarity and policy coordination under the European employment strategy." *European Union Politics* 7(2):167–89.

Franzese, Robert J. and Jude C. Hays. 2007. "Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data." *Political Analysis* 15(2):140–64.

Franzese, Robert J. and Jude C. Hays. 2008. "Interdependence in comparative politics: substance, theory, empirics, substance." *Comparative Politics Studies* 41(4-5):742–80.

Franzese, Robert J., Jude C. Hays and Aya Kachi. 2012. "Modeling History Dependence in Network-Behavior Coevolution." *Political Analysis* 20(2):175–190.

Franzese, Robert J., Jude C. Hays and Lena M. Schaffer. 2010. "Spatial, temporal, and spatiotemporal autoregressive probit models of interdependent binary outcomes: estimation, interpretation, and presentation." *American Political Science Association Annual Meeting* .

Freedom in the World 2007. 2007. *Washington, DC: Freedom House* .

Freeman, John R. and John E. Jackson. 2012. "Symposium on Models of Path Dependence ." *Political Analysis* 20(2):137–145.

Fukumoto, Kentaro. 2009. "What Happens Depends on When It Happens: Continuous or Ordered Event history Analysis Using Copula." Working Paper .

Gabizo, Mamoudou. 2005. "Foreign Aid and Democratization: Benina dn Niger Compared." *African Studies Review* 48:67–87.

Gasiorowski, M.J. 1996. "An overview of the political regime change dataset." *Comparative Political Studies* 29(4):469–83.

Geddes, Barbara. 1999. "What Do We Know about Democratization after Twenty Years?" *Annual Reviews of Political Science* 2:115–144.

Gisselquist, Rachel M. 2008. "Democratic Transition and Democratic Survival in Benin." *Democratization* 15:789–814.

Gleditsch, K.S. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46(5):712.

Gleditsch, K.S. and M.D. Ward. 2006. "Diffusion and the International Context of Democratization." *International Organization* 60(04):911–933.

Goodliffe, Jay and Darren Hawkins. 2011. "Dependence Networks and the Diffusion of Domestic Political Institutions." *Working Paper* .

Grzymala-Busse, Anna. 2006. "Authoritarian Determinants of Democratic Party Competition." *Party Politics* 12(3):415–437.

Gumbel, Emil J. 1959. "Multivariate Distributions with Given Margins." *Revista da Faculdade de Ciencias* 7(2):179–218.

Gumbel, Emil J. 1960. "Bivariate Exponential Distributions." *Journal of the American Statistical Association* pp. 698–707.

Hadenius, A. 1992. *Democracy and development*. Cambridge Univ Pr.

Hays, J.C. 2009. Bucking the System: Using Simulation Methods to Estimate and Analyze Systems of Equations with Qualitative and Limited Dependent Variables. In *2nd Annual St. Louis Area Methods Meeting (SLAMM), Washington University in St. Louis (https://netfiles. uiuc. edu/jchays/www/Hays_SLAMM09_Revised. pdf)*.

Hays, J.C. and A. Kachi. 2011. "Interdependent Duration Models in Political Science." *Working Paper* .

Hays, Jude C., Aya Kachi and Robert J. Franzese. 2010. "A spatial model incorporating dynamic, endogenous network interdependence: a political science application." *Statistical Methodology* 7(3):406–428.

Honoré, Bo. and Áureo de Paula. 2008. "Interdependent Durations." Working Paper .

Huntington, S.P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. University of Oklahoma Press.

Jackson, John E. and Ken Kollman. 2012. "Modeling, Measuring, and Distinguishing Path Dependence, Outcome Dependence, and Outcome Independence." *Political Analysis* 20(2):157–174.

Johnston, Jack and John DiNardo. 1997. *Econometric Methods. Fourth Edition*. McGraw-Hill.

King, Gary, James E. Alt, Nancy E. Burns and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American Journal of Political Science* 34(3):846–871.

Kopstein, Jeffrey S. and David A. Reilly. 2000. "Geographic Diffusion and the Transformation of the Postcommunist World." *World Politics* 53(1):1–37.

Laver, Michael. 1998. "Models of Government Formation." *Annual Review of Political Science* 1(1):1–25.

Laver, Michael. 2003. "Government Termination." *Annual Review of Political Science* 6(1):23–40.

Laver, Michael and Kenneth A. Shepsle. 1996. *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracies*. Cambridge University Press.

Lazarsfeld, P.F. and R.K. Merton. 1954. Friendship as a Social Process: A Substantive and Methodological Analysis. In *Freedom and Control in Modern Society*, ed. Berger M. Van Nostrand pp. 11–66.

Lazer, David. 2001. "The Co-Evolution of Individual and Network." *Journal of Mathematical Sociology* 25(1):69–108.

Lewis, Paul G. 2000. *Political Parties in Post-Communist Eastern Europe*. Routledge.

Lillard, Lee A. 1993. "Simultaneous Equations for Hazards." *Journal of Econometrics* 56:189–217.

Linz, Juan J. 1994. Presidential or Parliamentary Democracy: Does it Make a Difference? In *The Failure of Presidential Democracy*, ed. Juan J. Linz and Arturo Valenzuela. Johns Hopkins University Press.

Lipset, S.M. 1959. "Some social requisites of democracy: Economic development and political legitimacy." *The American Political Science Review* 53(1):69–105.

Lupia, Arthur and Kaare Strøm. 1995. "Coalition Termination and the Strategic Timing of Parliamentary Elections." *American Political Science Review* 89(3):648–665.

Mahoney, James. 2000. "Path Dependence in Historical Sociology." *Theory and Society* 29(4):507–548.

Mahoney, James. 2001. *The Legacies of Liberalism: Path Dependence and Political Regimes in Central America*. Johns Hopkins University Press.

Mainwaring, Scott and Anńbal Pérez-Liñán. 2008. "Regime Legacies and Democratization: Explaining Variance in the Level of Democracy in Latin America, 1978-2004." *Working Paper #354* .

Maltzman, Forrest and Charles R. Shipan. 2008. "Change, Continuity, and the Evolution of the Law." *American Journal of Political Science* 52(2):252–267.

Mansfield, E.D. and J. Snyder. 1995. "Democratization and the Danger of War." *International Security* 20(1):5–38.

Mansfield, E.D. and J. Snyder. 2007. "The Sequencing "Fallacy"." *Journal of Democracy* 18(3):5.

Mansfield, E.D. and J.L. Snyder. 2005. *Electing to Fight: Why Emerging Democracies Go to War*. The MIT Press.

Marshall, Monty G.., Keith Jaggers and Ted R. Gurr. 2006. N.d. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2004." . Forthcoming.

Martin, Lanny and Randy T. Stevenson. 2003. "Government Formation in Parliamentary Democracies." *American Journal of Political Science* 45(1):33–50.

McPherson, Miller., Lynn. Smith-Lovin and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." 27:415–444.

Mohrenberg, Steffen. N.d. "Analyzing Democratization with a Social Network Approach." *Working Paper*. Forthcoming.

Moore, Barrington. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Beacon Press.

Most, B.A. and H. Starr. 1990. "Theoretical and logical issues in the study of international diffusion." *Journal of Theoretical Politics* 2(4):391–412.

Murillo, M. Victoria and Andrew Schrank. 2005. "With a Little Help from My Friends: Partisan Politics, Transnational Alliances, and Labor Rights in Latin America." *Comparative Political Studies* 38(8):971–999.

Nelsen, R.B. 2006. *An Introduction to Copulas.* Springer Verlag.

Neumayer, Eric. 2008. "Distance, Power and Ideology: Diplomatic Representation in a World of Nation-States." *Area* 40(2):228–236.

O'Loughlin, J., L. Staeheli and E. Greenberg. 2004. *Globalization and its outcomes.* Guilford Press.

O'Loughlin, J., M.D. Ward, C.L. Lofdahl, J.S. Cohen, D.S. Brown, D. Reilly, K.S. Gleditsch and M. Shin. 1998. "The Diffusion of Democracy, 1946–1994." *Annals of the Association of American Geographers* 88(4):545–574.

Østerud, Øyvind. 2011. *Regime Transition and Virieties of Post-Communism.* Berliner Wissenschafts-Verlag chapter 3.

Page, Scott E. 2006. "Path Dependence." *Quarterly Journal of Political Science* 1(1):87–115.

Parks, Richard W. 1967. "Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated." *Journal of the American Statistical Association* pp. 500–509.

Pedigo, David. 2011. "Embassy Cables Reveal Brazil Supported Chile's Pinochet Regime." `http://www.santiagotimes.cl/chile/human-rights-a-law/23076-embassy-cables-reveal-brazil-supported-chiles-pinochet-regime`.

Pemstein, Daniel, Stephen Meserve and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.

Pierson, P. 2000. "Increasing returns, path dependence, and the study of politics." *American Political Science Review* 94(2):251–267.

Pierson, Paul. 2004. *Politics in Time: History, Institutions, and Social Analysis.* Princeton University Press.

Pierson, Paul and Theda Skocpol. 2002. Historical Institutionalism in Contemporary Political Science. In *Political Science: State of Discipline*, ed. Ira Katznelson and Helen V. Milner. W.W. Norton pp. 693–721.

Przeworski, A. and F. Limongi. 1997. "Modernization: Theories and facts." *World Politics* pp. 155–183.

Przeworski, A., M.E. Alvarez, J.A. Cheibub and F. Limongi. 2000. *Democracy and development.* Cambridge Univ. Press.

Putnam, R.D. 2002. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton University Press.

Quiroz Flores, Alejandro. 2008. "Copula Functions and Bivariate Distributions for Survival Analysis: An Application to Government Survival." NYU Department of Politics Working Paper .

Seely, Jennifer C. 2005. "The Legacies of Transition Governmetns: Post-Transition Dynamics in Benin and Togo." *Democratization* 12:357–377.

Shipan, Charles R. and Megan L. Shannon. 2003. "Delaying Justice (s): A Duration Analysis of Supreme Court Confirmations." *American Journal of Political Science* 47(4):654–668.

Simmons, B.A., F. Dobbin and G. Garrett. 2006. "Introduction: The international diffusion of liberalism." *International Organization* 60(04):781–810.

Simmons, Beth A. and Zachary Elkins. 2004. "The globalization of liberalization: Policy diffusion in the international political economy." *American Political Science Review* 98(01):171–189.

Snijders, Tom A.B. 1997. "Stochastic actor-oriented models for network change." *Evolution of social networks* pp. 149–172.

Snijders, Tom A.B. 2001. "The statistical evaluation of social network dynamics." *Sociological methodology* 31:361–395.

Snijders, Tom A.B. 2005. Models for longitudinal network data. In *Models and methods in social network analysis*, ed. Peter J. Carrington, John Scott and Stanley Wasserman. Cambridge Uninversity Press, New York chapter 11.

Snijders, Tom A.B., Christian Steglich and Michael Schweinberger. 2007. Modeling the co-evolution of networks and behavior. In *Longitudinal models in the behavioral and related sciences*, ed. K. van Montfort, H. Oud and A. Satorra. Lawrence Erlbaum, Mahwah, NJ chapter 3, pp. 41–71.

Starr, H. 1991. "Democratic dominoes: Diffusion approaches to the spread of democracy in the international system." *Journal of Conflict Resolution* pp. 356–381.

Steglich, Christian, Tom A.B. Snijders and Peter West. 2006. "Applying SIENA: an illustrative analysis of the co-evolution of adolescents' friendship networks, taste in music, and alcohol consumption." *Methodology: european journal of research methods for the behavioral and social sciences* 2(1):48–56.

Stinnett, Douglas M., Jaroslav Tir, Philip Schafer, Paul F. Diehl and Charles. Gochman. 2002. "The Correlates of War Project Direct Contiguity Data, Version 3." *Conflict Mangagement and Peace Science* 19(2):58–66.

Strøm, Kaare, Ian Budge and Michael J. Laver. 1994. "Constraints on Cabinet Formation in Parliamentary Democracies." *American Journal of Political Science* pp. 303–335.

Svolik, Milan. 2008. "Authoritarian Reversals and Democratic Consolidation." *American Political Science Review* 102(2):153–168.

Thelen, Kathleen. 1999. "Historical Institutionalism in Comparative Politics." *Annual Review of Political Science* 2:369–404.

Treier, S. and S. Jackman. 2008. "Democracy as a latent variable." *American Journal of Political Science* 52(1):201–217.

Vanhanen, T. 2003. *Democratization: A comparative analysis of 170 countries*. Routledge.

Waisman, Carols H. 1987. *Reversal of Development in Argentina*. Princeton University Press.

Warwick, Paul. and Stephen T. Easton. 1992. "The Cabinet Stability Controversy: New Perspectives on a Classic Poblem." *American Journal of Political Science* 36(1):122–146.

Warwick, Paul. V. 1992. "Economic Trends and Government Survival in West European Parliamentary Democracies." *American Political Science Review* 86(4):875–887.

Werner, Suzanne and Amy Yuen. 2005. "Making and Keeping Peace." *International Organization* 59(2):261–292.

Weyland, Kurt. 2009. "The Diffusion of Revolution: '1848' in Europe and Latin America." *International Organization* 63:391–423.

Wintrobe, R. 1998. *The Political Economy of Dictatorship*. Cambridge University Press.

World Development Indicators. N.d. World Bank.